EcoSal Plus
Cellular and Molecular Biology of
E. coli, Salmonella, and the *Enterobacteriaceae*

Reconstruction and Use of Microbial Metabolic Networks: the Core *Escherichia coli* Metabolic Model as an Educational Guide

JEFFREY D. ORTH, R. M. T. FLEMING, AND BERNHARD Ø. PALSSON
Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093–0412

ABSTRACT Biochemical network reconstructions have become popular tools in systems biology. Metabolicnetwork reconstructions are biochemically, genetically, and genomically (BiGG) structured databases of biochemical reactions and metabolites. They contain information such as exact reaction stoichiometry, reaction reversibility, and the relationships between genes, proteins, and reactions. Network reconstructions have been used extensively to study the phenotypic behavior of wild-type and mutant stains under a variety of conditions, linking genotypes with phenotypes. Such phenotypic simulations have allowed for the prediction of growth after genetic manipulations, prediction of growth phenotypes after adaptive evolution, and prediction of essential genes. Additionally, because network reconstructions are organism specific, they can be used to understand differences between organisms of species in a functional context. There are different types of reconstructions representing various types of biological networks (metabolic, regulatory, transcription/translation). This chapter serves as an introduction to metabolic and regulatory network reconstructions and models and gives a complete description of the core *Escherichia coli* metabolic model. This model can be analyzed in any computational format (such as MATLAB or Mathematica) based on the information given in this chapter. The core *E. coli* model is a small-scale model that can be used for educational purposes. It is meant to be used by senior undergraduate and first-year graduate students learning about constraint-based modeling and systems biology. This model has enough reactions and pathways to enable interesting and insightful calculations, but it is also simple enough that the results of such calculations can be understood easily.

Received: 3 June 2009
Accepted: 8 September 2009
Posted: 1 February 2010
Supercedes previous posting at EcoSal.org.
Editor: Peter D. Karp, SRI International, Menlo Park, CA
Citation: EcoSal Plus 2013; doi:10.1128/ecosalplus.10.2.1.
Correspondence: Bernhard Ø. Palsson: palsson@ucsd.edu

doi:10.1128/ecosalplus.10.2.1

INTRODUCTION

Biochemical network reconstructions have become popular tools in systems biology (1). There are different types of reconstructions representing various types of biological networks (metabolic, regulatory, transcription/translation), although metabolic network reconstructions have been used the most extensively. Early reconstructions of *Escherichia coli* metabolism were small, containing only central metabolic reactions (2). Today, there are many genome-scale reconstructions available, based on all known metabolic genes

in the annotated genome of an organism along with other data sources (3, 4, 5, 6). Metabolic network reconstructions are biochemically, genetically, and genomically (BiGG) structured databases of biochemical reactions and metabolites. They contain information such as exact reaction stoichiometry, reaction reversibility, and the relationships between genes, proteins, and reactions. Reconstructed networks serve as flexible BiGG knowledge bases (7), storing curated information in a useful format while allowing for content to be updated based on new research. Although many organisms have similar central metabolic networks, there can be differences even between two closely related organisms. Network reconstructions are therefore organism specific (8). For most applications of network reconstructions, it is necessary to convert the network into a mathematical model. Metabolic models are usually formulated as a stoichiometric matrix, while regulatory network models are often formulated as Boolean networks.

There have been many practical uses of network reconstructions. Some reconstructions have been used as tools to study bacterial evolution. The effects on metabolism of adding or removing genes from the network can be simulated, enabling studies of horizontal gene transfer (9), adaptation to new environments (10), and evolution to minimal genomes (11). Reconstructions can also be used for analysis of network properties. In these studies, methods have been developed to determine the interactions between different sets of reactions and compounds, improving our understanding of the organisms under investigation. Some examples include identification of alternate optimal network states (12), identification of sets of coupled reactions (13), and studies of the states of regulatory networks (14, 15). It has also been determined by simulating thousands of growth conditions that *E. coli* contains a set of common, high-flux backbone reactions (16). Network reconstructions have been extensively used to study the phenotypic behavior of wild-type and mutant stains under a variety of conditions, linking genotypes with phenotypes. These predictions have been verified by experimental studies (17). Such phenotypic simulations have allowed for the prediction of growth after genetic manipulations (18, 19), prediction of growth phenotypes after adaptive evolution (20), and prediction of essential genes (21). Another promising use of reconstructions is in the discovery of unknown biological features. By comparing experimental data such as growth phenotypes (22), metabolic flux measurements (23), or gene essentiality (24) with model-based predictions, missing content in reconstructions can be identified. The reconstructions can be updated and new biological knowledge can be elucidated. Finally, network reconstructions have proven to be very useful for metabolic engineering and synthetic biology (25). Because of the capacity of models to be used to predict growth and metabolite secretion phenotypes, it is possible to predict the genetic interventions most likely to produce a strain with the desired properties (26). Model-based algorithms can even predict nonintuitive designs that couple production of desired metabolites to cell growth (27, 28).

This chapter serves as an introduction to metabolic and regulatory network reconstructions and models, and gives a complete description of the core *E. coli* metabolic model. This model can be encoded and analyzed in any computational format (such as MATLAB or Mathematica) based on the information given in this chapter. The core *E. coli* model is a small-scale model that can be used for educational purposes. It is meant to be used by senior undergraduate and first-year graduate students learning about constraint-based modeling and systems biology. This model has enough reactions and pathways to enable interesting and insightful calculations, but it is also simple enough that the results of such calculations can be easily understood. This model is also useful for testing and evaluating new constraint-based analysis methods, since its small scope makes troubleshooting and interpretation of results easier.

THE FOUR-STEP PROCESS OF BUILDING GENOME-SCALE METABOLIC RECONSTRUCTIONS

The construction of a genome-scale metabolic network is a long-term process (from many months to several years, depending on the size of the network) consisting of four major steps, each step requiring the use of different types of biological data (Fig. 1). In the first step, an organism's annotated genome is used to generate a draft reconstruction. Second, the draft reconstruction is curated in a long process that involves the study of many highly specific data sources. In the third step, the reconstruction is converted to a mathematical model, and model-based simulations can be compared with phenotypic data. In the fourth and final step, high-throughput data can be integrated with the model, allowing for biological discovery and iterative model refinement. After the model is complete, it can be used for a variety of purposes (see "Uses of Metabolic Models," below).

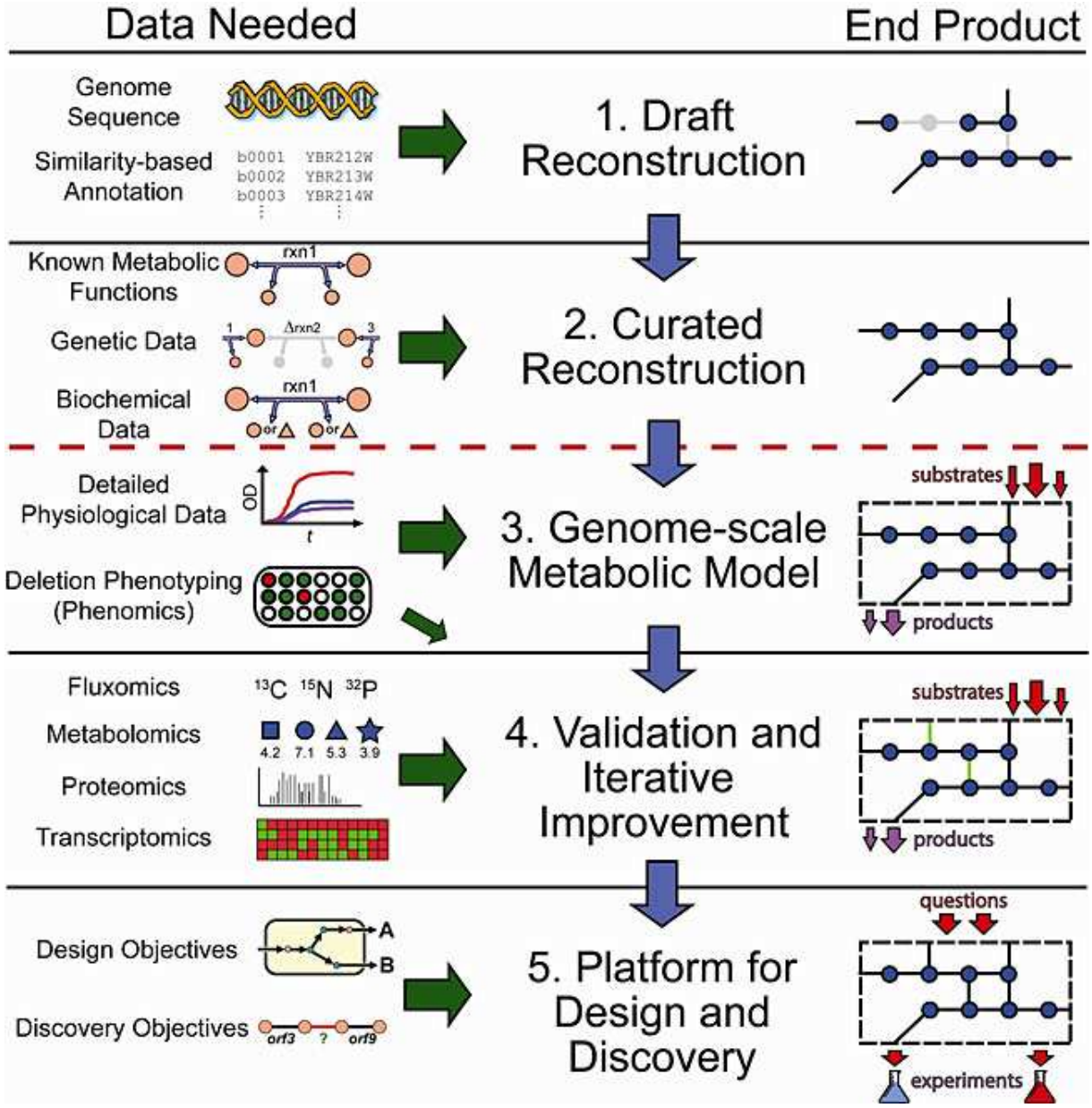

Figure 1 The phases and data utilized in generating a metabolic reconstruction. The genome-scale metabolic reconstruction process can be summarized in four major phases, each of the latter phases building off the previous one. The fifth phase is use of the complete reconstruction for practical purposes (see "Uses of Metabolic Models"). Characteristic of the reconstruction process is the iterative refinement of reconstruction content that is driven by experimental data and occurs in phases 2 to 4. For each phase, specific data types are necessary and these range from high-throughput data types, e.g., metabolomics, to detailed studies characterizing individual components, e.g., biochemical data for a particular reaction. For example, the genome annotation can provide a parts list of a cell, while genetic data can provide information about the contribution of each gene product toward a phenotype when removed or mutated. The product generated from each reconstruction phase can be utilized and applied to examine a growing number of questions with the final product having the broadest applications. Figure adapted from reference 7.

Initial Reconstruction Based on an Annotated Genome

The first step in building any new genome-scale reconstruction, whether of metabolic or regulatory networks, is to use the genome annotation for the desired organism to generate an initial list of functions. Genomes have been annotated to some degree for hundreds of microbial organisms, and they provide several types of valuable information. First, they contain the genome sequence of an organism, from which open reading frames (ORFs) can be identified. The function of each ORF can then be determined through a variety of methods (22). The

strongest evidence for the function of a particular ORF usually comes from direct biochemical analysis, such as isolation and characterization of the function of an enzyme. *E. coli* is a very well studied model organism and many of its ORFs have been experimentally characterized (29). Unfortunately, for many other organisms, few biochemical data are available. To identify the ORFs in these organisms (and also to identify uncharacterized ORFs in well studied organisms), their genomic sequences are compared with the genomes of other organisms to identify homologous genes. In silico methods can also be used for annotation, including methods that identify genes based on protein-protein interactions, transcriptomics, phylogenetic profiles, protein fusion, and operon clustering (30). These methods typically allow for 40 to 70% of the genes in a new genome to be annotated. When a high-quality genome annotation is not available for a particular organism, it becomes more challenging to build a reconstruction of that organism on a genome scale.

There are organism-specific databases for some genome annotations, including EcoCyc (29) for *E. coli* and the *Saccharomyces* Genome Database (SGD) (31) and Comprehensive Yeast Genome Database (CYGD) (32) for yeast. For many other microbial organisms (33), Comprehensive Microbial Resource (CMR) (34), Genome Reviews (35), and Integrated Microbial Genomes (IMG) (36) contain useful genomic information. All of the metabolic genes identified in the genome annotation for the desired organism can be assembled into an initial parts list. From this initial list of genes, an initial list of enzymes and the reactions they catalyze can be constructed by mapping each gene to one or more reactions according to information from a database. The data used for this mapping can be included in the genome annotations or it can be obtained from metabolic databases such as KEGG (37), BRENDA (38), ENZYME (39), MetaCyc (40), the SEED (41), or TransportDB (42). Most databases include EC (Enzyme Commission) numbers that can be used to easily identify the enzymes and reactions associated with a particular gene. With the appropriate databases, the reactions known to be associated with each gene can be identified. Some of these databases will be more useful for certain organisms than others. The process of building an initial reconstruction from a genome annotation and reaction information from databases can be performed manually, or it can be partially or fully automated. Tools available for building draft reconstructions include SimPheny (Genomatica, Inc., San Diego, CA), PathoLogic (43), and PRIAM (44). SimPheny is a commercial software platform for building and analyzing metabolic constraint-based models. It can download the annotated genome of an organism and provide a framework for manually associating metabolic genes with reactions. SimPheny is also useful in the other stages of building reconstructions because it contains tools for manual curation and model quality control and quality assurance, as well as tools for performing simulations and analyzing experimental data. PathoLogic, part of the Pathway Tools software system, is a tool for mapping genes to reactions in an automated manner. It requires a fully annotated genome, and uses EC numbers, Gene Ontology terms (45), or the annotated gene names to predict which reactions are associated with a particular gene. It can then predict which pathways are present in an organism by comparing the predicted set of reactions with a reference database such as MetaCyc. PRIAM is another automated method that identifies enzymes in any genome sequence. This program uses all of the known sequences for any individual enzyme in the ENZYME database to identify the characteristic sequence modules of that enzyme. Specific rules that can identify an enzyme based on which modules are present in a sequence are then formulated. PRIAM forms these modules and rules for every enzyme in the database, and then uses scoring matrices to identify modules in the genome of interest. It then uses the rules to predict which enzyme is associated with every gene in the genome. This algorithm can be very useful because it does not require a fully annotated genome. The result of the initial mapping process is a draft reconstruction that lists most of the metabolic genes and reactions in an organism with reasonable accuracy.

Curation of the Initial Reconstruction

The next step in the reconstruction process is to manually curate the initial reconstruction. Any reconstruction resulting from a fully automated procedure will be incomplete and, in some cases, incorrect. Some reactions will be missing because it is not known which genes encode their enzymes, leaving gaps in pathways. Other reactions may be mistakenly included because of incorrect genome annotation or nonspecific information in databases. Reactions in the reconstruction may have incorrect or unbalanced stoichiometries or cofactor usage, because these attributes are often unique to enzymes in specific organisms (8). Gene-protein-reaction associations (GPRs) must also be included to formally connect

reactions to one or more functional proteins, and every protein to one or more known genes (46). To correct any mistakes and improve the reconstruction, a researcher must manually curate the list of reactions by using data from many different sources. Organism-specific textbooks and databases are very useful for this purpose. The genome-scale *E. coli* reconstruction *i*AF1260 (4) relied heavily on the textbook Escherichia coli *and* Salmonella: *Cellular and Molecular Biology* (47). Unfortunately, such texts are usually not available for less-well-studied organisms. Literature data, from both primary and review articles, are also extremely useful. These articles can contain useful and specific information about reaction stoichiometry and directionality and can indicate the presence of many reactions with unknown genes. Many different types of studies can be useful, including enzyme assays, gene knockout studies, metabolomic studies including flux measurements, and protein localization studies. The information from these sources often cannot be found in databases. The manual curation process is extremely labor intensive, usually requiring the study of hundreds of literature sources over a period of several months or years.

Conversion of the Reconstruction to a Computational Model

After a high-quality reconstruction has been assembled, it must be converted into a genome-scale constraint-based model to be further analyzed (1). A reconstruction is a BiGG knowledge base, a list of stoichiometrically balanced reactions and their associated genes and proteins. A model is a network in a mathematical format with defined system boundaries and constraints on the reactions (2). While a reconstruction is unique to an organism, many different models (e.g., condition specific) can be derived from a reconstruction. Metabolic network models are usually encoded in a stoichiometric matrix, in which each unique metabolite is represented by a row in the matrix, and each reaction is represented by a column. The entries in each column are the stoichiometric coefficients of the metabolites in a reaction, with negative coefficients for consumed metabolites and positive coefficients for produced metabolites. The properties of this matrix can be investigated through various constraint-based analysis methods, including flux balance analysis (2, 48, 49, 50). This method uses linear optimization to identify optimal reaction flux distributions of the network given a set of minimum and maximum reaction rates and an objective, such as maximum cellular growth. To simulate growth with a genome-scale metabolic model, a biomass reaction is needed. This is an organism-specific reaction that drains specified metabolites from the network, representing the metabolic precursors that contribute to biomass (2, 49). To construct a biomass reaction, the relative amounts of nucleic acids, lipids, proteins, and other macromolecules of an organism must be known. These macromolecules can then be broken down into building blocks such as nucleotides or amino acids. The relative amounts of each of these building blocks form the stoichiometric coefficients of the biomass reaction (see "Biomass Reaction," below). Experimentally determined growth data must also be used to determine the amount of energy needed for growth and for non-growth-associated maintenance functions, representing the energy demands of the cell (4). Once the biomass reaction has been constructed, flux balance analysis (2, 49) can be used to predict optimal growth rates under many different conditions (50). A completed genome-scale metabolic model can be used to assess the quality of the reconstruction. There must be continuous pathways to every biomass precursor or the in silico cell cannot grow. Other gaps in the network can be identified by unreachable reactions or metabolites (51). Another common test of a new model is to compare growth simulations under different conditions with actual growth data (17).

Model Validation and Iterative Improvement

Genome-scale metabolic models can be used to map many types of experimental data to a biological network, allowing for the integration of different data types. These data can be compared with model predictions, and the discrepancies can lead to discovery of new reactions and pathways. High-throughput screens for growth on different media conditions can be used to reveal previously unknown substrate uptake pathways. In vivo knockout screens and synthetic lethal screens can be compared with in silico-predicted knockout effects (Fig. 2), with discrepancies indicating the presence of alternative metabolic pathways (22). Metabolomic data can predict the presence of metabolites not accounted for in a reconstruction, necessitating the addition of new production and utilization pathways. Proteomic and transcriptomic data can be used to suggest the genes and proteins that fill gaps in a reconstruction (52). As new biological features and capabilities are discovered, the model can be improved by incorporating these new data. The updated model can then be used to probe different pathways and

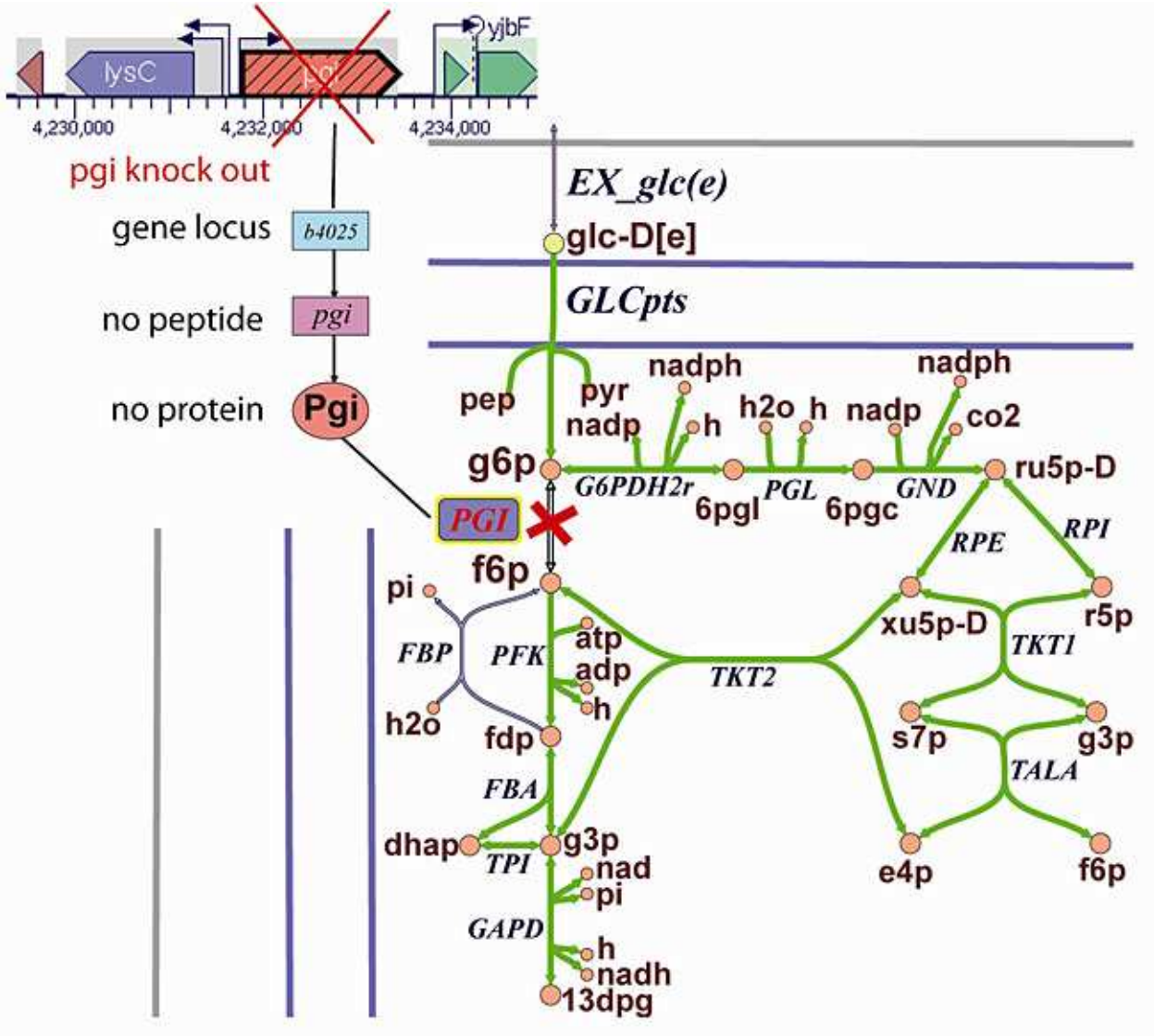

Figure 2 An illustration of an in silico knockout of glucose-6-phosphate isomerase, *PGI* (53), which catalyzes the conversion of D-glucose 6-phosphate, g6p, to D-fructose 6-phosphate, f6p. The Δ*pgi* in silico model predicts that loss of glucose-6-phosphate isomerase results in use of the pentose phosphate pathway as the primary route of glucose catabolism, as has been observed experimentally (54).

features, leading to an iterative cycle of discovery and model improvement.

BUILDING GENOME-SCALE REGULATORY NETWORK RECONSTRUCTIONS

The process of reconstructing transcriptional regulatory networks is not as well developed as the process for building metabolic reconstructions. To date, only a few examples of genome-scale regulatory reconstructions exist (55, 56, 57). The oldest of these is *i*MC1010, which contains the regulators of *E. coli* metabolism (58). This model was built by a process similar to that used for metabolic reconstructions, relying on a variety of data types including the genome annotation and literature sources. A gene expression study was then conducted in which gene expression under aerobic and anaerobic conditions was compared in several different transcription factor knockout strains. By analyzing the discrepancies between the experimental results and the model predictions, the model was iteratively updated and improved.

Several automated methods for inferring transcriptional regulatory networks have been developed recently (55, 59, 60, 61). Progress in high-throughput experimental

methods is allowing for transcriptional regulatory reconstructions to be assembled in a top-down, automated manner (62, 63, 64). The connectivity of an organism's transcriptional regulatory network can be determined by performing ChIP-chip experiments (chromatin immunoprecipitation followed by microarray hybridization). In these studies, all of the DNA binding sites of a particular transcription factor on an entire genome under a particular set of conditions can be identified in vivo. First, proteins are fixed to genomic DNA in living cells, and then the DNA is extracted and sheared into small fragments. Next, the fragments with a particular transcription factor bound are filtered out by using antibodies, and the fragments are identified by hybridization to a DNA microarray. When ChIP-chip experiments are repeated under a variety of conditions, all of the binding sites of a transcription factor can be found, identifying all of the genes regulated by that transcription factor.

A set of ChIP-chip experiments must be run for every transcription factor to elucidate an entire regulatory network. As useful as these experiments are, they do not reveal what the effect of a transcription factor on its targets is, or how different transcription factors interact. The direct and indirect effects of transcription factors can be determined by expression profiling of strains with those transcription factor genes knocked out (65). Performing so many high-throughput experiments is a very expensive and time-consuming process, but improvements in parallel sequencing technologies should improve the effectiveness of these approaches. ChIP-seq experiments (chromatin immunoprecipitation followed by DNA sequencing) are an example of this trend (62, 63, 66, 67, 68).

As with metabolic reconstructions, regulatory reconstructions must be converted to computational models to utilize their predictive potential. Several different model types can be used. The most common modeling framework is the Boolean model, which represents the connections between genes or other variables as logical rules. Boolean models can qualitatively describe the functions of a regulatory network and make accurate predictions of behavior (69). An equivalent structure to the Boolean model is the "regulation matrix" (70). By reformulating a Boolean model as a matrix, more advanced mathematical analysis is possible, and all of the possible expression states of the model can be sampled (71). Newer regulatory network models are being formulated in structures similar to the stoichiometric matrix, allowing them to be interrogated with constraint-based analysis methods (72).

DESCRIPTION OF THE CORE *E. COLI* METABOLIC RECONSTRUCTION

Here, the biochemistry of the reactions in the core *E. coli* metabolic model is summarized. An overview of the reactions is given in Fig. 3. Where possible, reactions are described in the context of their major functional pathway. Reactions occasionally participate in more than one pathway and, in such cases, this is highlighted. The *E. coli* core model is based on the first stoichiometric reconstruction of *E. coli* fueling pathways (2). The current model contains the reactions of glycolysis, the pentose phosphate shunt, the tricarboxylic acid cycle, the glyoxylate cycle, gluconeogenesis, anaplerotic reactions, the electron transport chain and oxidative phosphorylation, the transfer of reducing equivalents, fermentation, and nitrogen metabolism.

In this model, metabolites and reactions are given both full names and abbreviations. Metabolite abbreviations are lowercase, and extracellular metabolites are denoted with the suffix "[e]," e.g., extracellular acetate is abbreviated "ac[e]". This reconstruction does not distinguish between the periplasmic space and the extracellular medium. All metabolites that are not denoted as extracellular are cytosolic. In many reconstructions, cytosolic metabolites use the suffix "[c]," but here, this is omitted for clarity. In the figures describing the metabolic network, such as Fig. 3, cytosolic metabolites are represented by orange circles and extracellular metabolites are represented by yellow circles. Reaction abbreviations are uppercase and italicized. For example, acetaldehyde dehydrogenase (acetylating) is abbreviated as *ACALD*. There are several common suffixes used in the reaction abbreviations, including "abc" (ATP-binding cassette transporter), "i" (irreversible), "r" (reversible), and "t" (transport). Most reactions in the reconstruction are named after the enzymes that catalyze them. In this section, the text often uses the reaction abbreviations to refer to the enzymes catalyzing the reactions.

In the figures, metabolic reactions are represented as blue arrows, and the reaction abbreviations are inside blue boxes with yellow outlines. Certain reactions are assumed to be effectively irreversible when thermodynamic considerations are taken into account (4, 73). In brief, if the in vivo change in Gibbs energy for a biochemical reaction

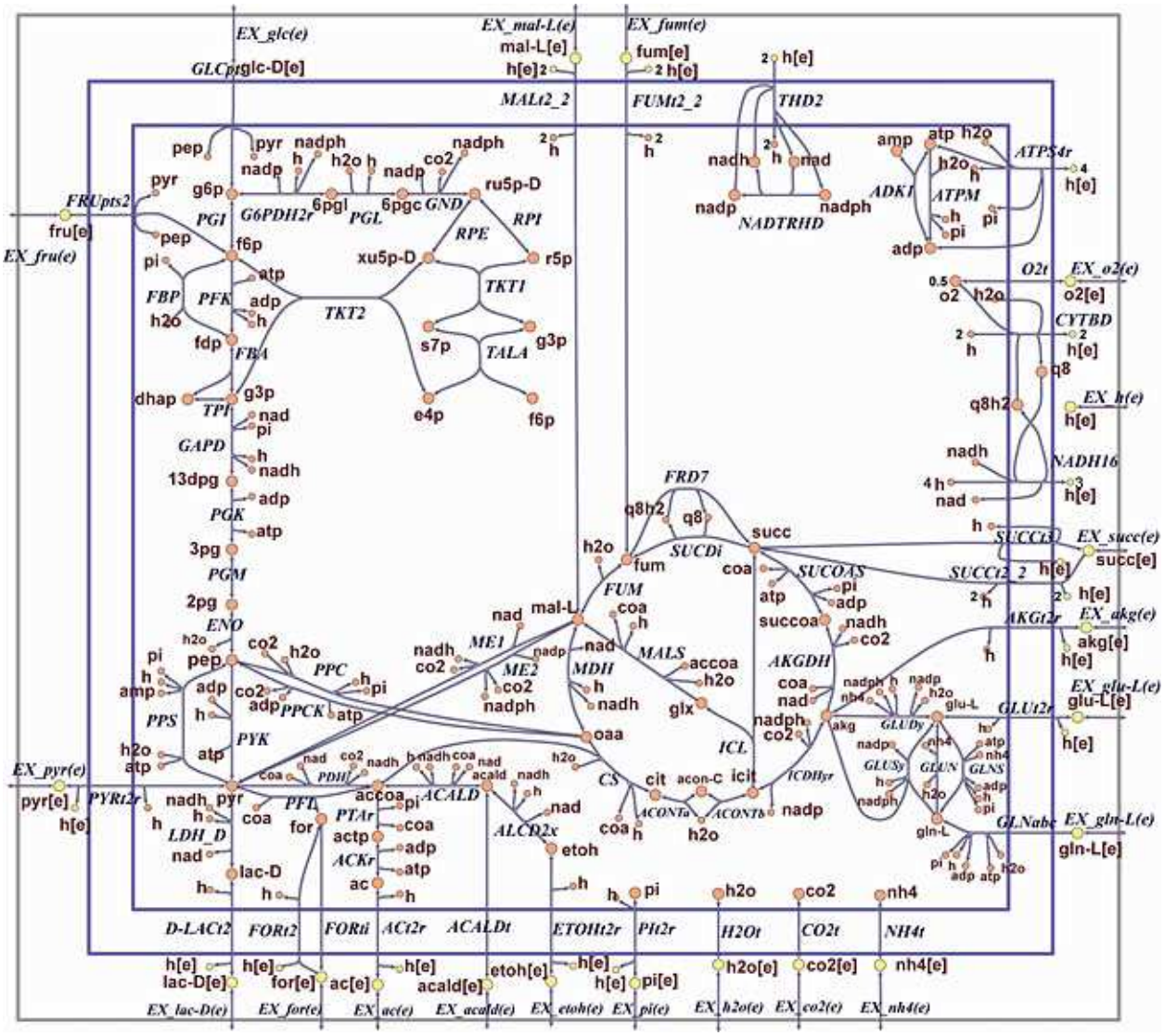

Figure 3 An overview of core *E. coli* metabolism. The outer gray box represents the boundary between the model and the environmental source of substrates and sink for waste metabolites. The outer and inner blue boxes represent the outer and inner surface of the cytoplasmic membrane. The periplasmic space is outside the scope of the core model. Cytosolic metabolites are represented by orange circles and extracellular metabolites are represented by yellow circles.

is highly negative, then it is assumed that net flux is always in the forward direction. In all figures, effectively irreversible reactions are denoted with arrowheads in one direction only. The tables associated with each pathway also indicate reactions that are effectively irreversible with the symbol → in the reaction equation. Reversible reactions use the symbol ↔ in the reaction equation.

GPRs are shown adjacent to each reaction. The reconstruction associates each reaction with an enzyme or an enzyme complex. Some reactions may be catalyzed by more than one enzyme and this distinction is also represented (Fig. 4). A few reactions are known to occur in *E. coli*, but the corresponding gene has yet to be identified. These orphan reactions are discussed further in "Discovery." Each protein is associated with a gene name and genomic locus. The genomic locus ties a reaction to one particular unique nucleotide sequence in the 4,639,675-base-pair *E. coli* K-12 MG1655 circular chromosome (74). The functional genomic structure, such as operon structure or transcriptional units, is not specifically represented in the *E. coli* core reconstruction but is represented in more comprehensive models of *E. coli* (72). The charge of each metabolite is included in the core model to determine the proper chemical formula

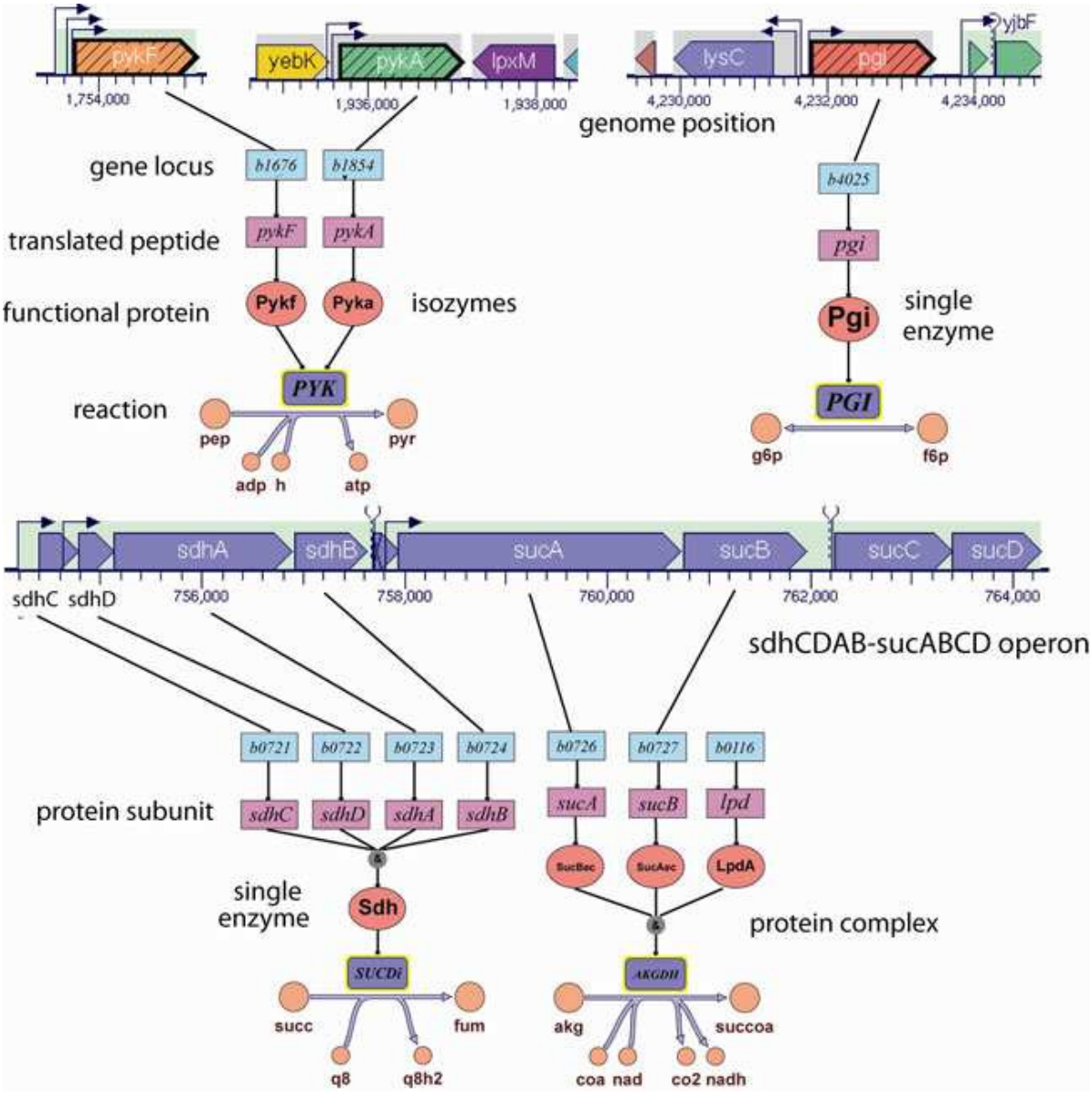

Figure 4 An example of GPRs from the *E. coli* core model. Genes are represented by blue boxes and designated by their locus name, translated peptides are represented by purple boxes, functional proteins are represented by red ovals, and reactions are labeled with blue boxes. For isozymes, two different proteins are connected to the same reaction. For proteins with multiple peptide subunits, the peptides are connected with an "&" above the protein. For complexes of multiple functional proteins, the proteins are connected with an "&" above the reaction. The genomic context of some of these genes is highlighted. Certain genes for the same reaction, e.g., *pykF* and *pykA*, are encoded by genes in operons widely separated on the genome. Operons are represented by shaded rectangles around one or more genes. Genes are represented by rectangles with one side pointed to denote the direction of the sense strand. Other operons contain multiple genes that encode protein subunits in a large protein. In this case, the same *sdhCDAB-sucABCD* operon that codes for the *SUCDi* proteins (75) also codes for two proteins of the 2-oxoglutarate dehydrogenase enzyme complex, *AKGDH*. Genome context figures created by use of the Pathway Tools Genome Browser from EcoCyc (29, 76).

in *E. coli*. These charges were determined by using the pK_a of each metabolite at a pH of 7.2 (4).

The reactions and pathways in the core model were chosen to represent the most well known and widely studied metabolic pathways of *E. coli*. These pathways are often the subjects of textbook chapters and should be familiar to most readers with a basic biochemistry background. As much as possible, the reactions and GPRs were taken directly from the *i*AF160 genome-scale *E. coli* reconstruction (4). Some pathways, such as the electron transport system, were greatly simplified to limit the

scope of the model and ensure that every reaction is understandable. This model contains a total of 72 metabolites and 95 reactions. There are 20 extracellular metabolites and 52 intracellular metabolites, with a total of 54 unique metabolites (most extracellular metabolites are just extracellular versions of intracellular metabolites). There are 20 exchange reactions, one for each extracellular metabolite. The model also has 25 transport reactions, 49 metabolic reactions, and one biomass reaction (see "Biomass Reaction," below).

Glycolysis

One of the most important and widely studied metabolic pathways is glycolysis, a series of ten chemical reactions that convert one 6-carbon glucose molecule into two 3-carbon pyruvate molecules (see Table 1, Table 2, and Fig. 5). In these reactions, a net two molecules of adenosine triphosphate (ATP) are produced by substrate-level phosphorylation, and two molecules of reduced nicotinamide adenine dinucleotide (NADH) are also produced. In addition to producing these energy and redox carriers, glycolysis also produces several compounds that are precursors for *E. coli* biomass (see "Biomass reaction," below).

Table 1 Glycolysis reactions

Abbr.	Reaction	Equation
GLCpts	D-Glucose transport via PEP:Pyr PTS	glc-D[e] + pep → g6p + pyr
PGI	Glucose-6-phosphate isomerase	g6p ⟷ f6p
FRUpts2	Fructose transport via PEP:Pyr PTS (f6p generating)	fru[e] + pep → f6p + pyr
PFK	Phosphofructokinase	atp + f6p → adp + fdp + h
FBP	Fructose-bisphosphatase	fdp + h2o → f6p + pi
FBA	Fructose-bisphosphate aldolase	fdp ⟷ dhap + g3p
TPI	Triose-phosphate isomerase	dhap ⟷ g3p
GAPD	Glyceraldehyde-3-phosphate dehydrogenase	g3p + nad + pi ⟷ 13dpg + h + nadh
PGK	Phosphoglycerate kinase	3pg + atp ⟷ 13dpg + adp
PGM	Phosphoglycerate mutase	2pg ⟷ 3pg
ENO	Enolase	2pg ⟷ h2o + pep
PYK	Pyruvate kinase	adp + h + pep → atp + pyr
PPS	Phosphoenolpyruvate synthase	atp + h2o + pyr → amp + 2 h + pep + pi

Table 2 Glycolysis metabolites

Abbr.	Metabolite	Formula	Charge
glc-D	D-Glucose	$C_6H_{12}O_6$	0
g6p	D-Glucose 6-phosphate	$C_6H_{11}O_9P$	−2
fru	D-Fructose	$C_6H_{12}O_6$	0
f6p	D-Fructose 6-phosphate	$C_6H_{11}O_9P$	−2
fdp	D-Fructose 1,6-bisphosphate	$C_6H_{10}O_{12}P_2$	−4
dhap	Dihydroxyacetone phosphate	$C_3H_5O_6P$	−2
g3p	Glyceraldehyde 3-phosphate	$C_3H_5O_6P$	−2
13dpg	3-Phospho-D-glyceroyl-phosphate	$C_3H_4O_{10}P_2$	−4
3pg	3-Phospho-D-glycerate	$C_3H_4O_7P$	−3
2pg	D-Glycerate-2-phosphate	$C_3H_4O_7P$	−3
pep	Phosphoenolpyruvate	$C_3H_2O_6P$	−3
pyr	Pyruvate	$C_3H_3O_3$	−1
h	H^+	H	1
h2o	H_2O	H_2O	0
amp	Adenosine monophosphate	$C_{10}H_{12}N_5O_7P$	−2
adp	Adenosine diphosphate	$C_{10}H_{12}N_5O_{10}P_2$	−3
atp	Adenosine triphosphate	$C_{10}H_{12}N_4O_{13}P_3$	−4
pi	Phosphate	HO_4P	−2
nad	Nicotinamide adenine dinucleotide (NAD^+)	$C_{21}H_{26}N_7O_{14}P_2$	−1
nadh	Nicotinamide adenine dinucleotide-reduced	$C_{21}H_{27}N_7O_{14}P_2$	−2

Glycolysis begins at the phosphoenolpyruvate:pyruvate phosphotransferase protein complex, which actively translocates hexoses across the inner cytoplasmic membrane (77). Certain proteins of the phosphoenolpyruvate: pyruvate phosphotransferase complex (PEP:Pyr PTS) are carbohydrate specific, but in each case, transport is driven by transfer of the phosphate group of phosphoenolpyruvate to the carbohydrate. In the reaction D-glucose transport via PEP:Pyr PTS, *GLCpts*, phosphoenolpyruvate donates the phosphate group to glucose to form D-glucose 6-phosphate, and the dephosphorylated remainder of phosphoenolpyruvate is pyruvate (78). The same general procedure generates D-fructose 6-phosphate from fructose in the reaction fructose transport via PEP:Pyr PTS (f6p generating), *FRUpts2* (79). The interconversion of D-glucose 6-phosphate and D-fructose 6-phosphate is catalyzed by glucose-6-phosphate isomerase, *PGI* (53). Phosphofructokinase, *PFK*, catalyzes the transfer of a phosphate group from ATP to D-fructose 6-phosphate to form D-fructose 1,6-bisphosphate and adenosine diphosphate (ADP) (80, 81). This reaction is

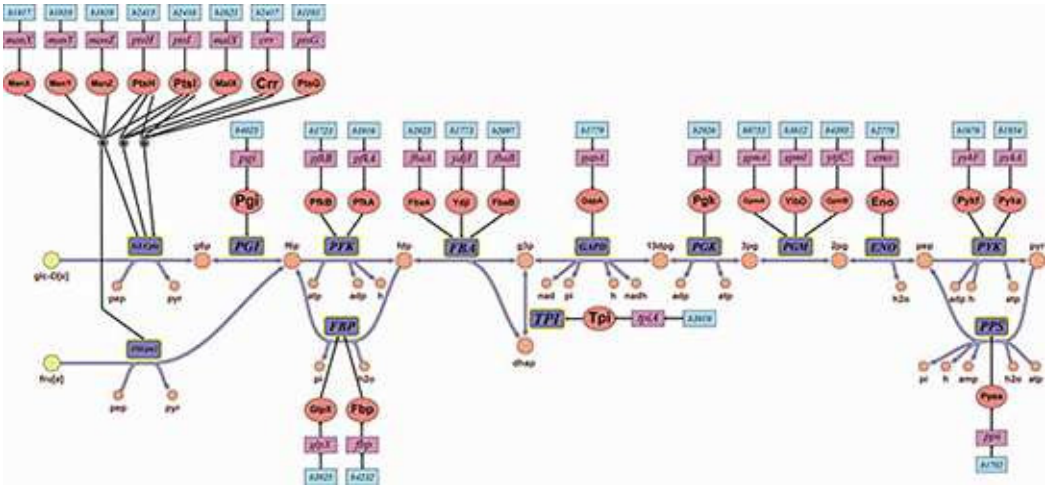

Figure 5 Glycolysis consists of a sequence of reactions that converts sugars, such as glucose or fructose, into a series of precursors for biomass components, terminating with pyruvate. There is also the concomitant production of relatively small amounts of ATP and NADH. Gluconeogenesis refers to the production of sugars by reversing the flux through thermodynamically reversible reactions in the glycolytic pathway while bypassing reactions that are effectively thermodynamically irreversible.

effectively thermodynamically irreversible. However, in the reverse direction, the dephosphorylation of D-fructose 1,6-bisphosphate to form D-fructose 6-phosphate is catalyzed by fructose-bisphosphatase, *FBP* (82). The role of this reaction is discussed further in "Glyoxylate Cycle, Gluconeogenesis, and Anaplerotic Reactions," below.

Fructose-bisphosphate aldolase, *FBA*, splits the 6-carbon D-fructose 1,6-bisphosphate into two 3-carbon molecules, dihydroxyacetone phosphate and glyceraldehyde 3-phosphate (82, 83, 84, 85). Triose-phosphate isomerase, *TPI*, rapidly and reversibly structurally rearranges dihydroxyacetone phosphate to glyceraldehyde 3-phosphate (86). A linear sequence of four reversible reactions catalyzed by glyceraldehyde-3-phosphate dehydrogenase, *GAPD* (87), phosphoglycerate kinase, *PGK* (88), phosphoglycerate mutase, *PGM* (89), and enolase, *ENO* (90), converts glyceraldehyde 3-phosphate to phosphoenolpyruvate. This sequence of reactions also reduces one NAD^+ to form NADH in the glyceraldehyde-3-phosphate dehydrogenase reaction and yields one high-energy currency metabolite, ATP, in the phosphoglycerate kinase reaction. In the final step of glycolysis, pyruvate kinase, *PYK*, catalyzes the transfer of the phosphate group of phosphoenolpyruvate to ADP resulting in the production of pyruvate and ATP (91, 92). A cursory glance at Fig. 3 reveals that pyruvate is an important precursor involved in many pathways. It can be converted into acetyl-CoA, which provides carbon for the tricarboxylic acid cycle, as discussed in "Tricarboxylic Acid Cycle." It can also be converted into lactate as part of fermentation, discussed in "Fermentation."

Pentose Phosphate Shunt

The primary function of the pentose phosphate shunt is to provide the 5-carbon and 4-carbon biosynthetic precursors α-D-ribose 5-phosphate and D-erythrose 4-phosphate. α-D-Ribose 5-phosphate and D-erythrose 4-phosphate can be produced by either of two parallel pathways, the decarboxylating oxidative pathway or the nonoxidative pathway (see Table 3, Table 4, and Fig. 6).

Table 3 Pentose phosphate shunt reactions

Abbr.	Reaction	Equation
G6PDH2r	Glucose-6-phosphate dehydrogenase	g6p + nadp ⟷ 6pgl + h + nadph
PGL	6-Phosphogluconolactonase	6pgl + h2o → 6pgc + h
GND	Phosphogluconate dehydrogenase	6pgc + nadp → co2 + nadph + ru5p-D
RPI	Ribose-5-phosphate isomerase	r5p ⟷ ru5p-D
TKT2	Transketolase	e4p + xu5p-D ⟷ f6p + g3p
TALA	Transaldolase	g3p + s7p ⟷ e4p + f6p
TKT1	Transketolase	r5p + xu5p-D ⟷ g3p + s7p
RPE	Ribulose-5-phosphate 3-epimerase	ru5p-D ⟷ xu5p-D

Table 4 Pentose phosphate metabolites

Abbr.	Metabolite	Formula	Charge
6pgl	6-Phospho-D-glucono-1–5-lactone	$C_6H_9O_9P$	−2
6pgc	6-Phospho-D-gluconate	$C_6H_{10}O_{10}P$	−3
ru5p-D	D-Ribulose 5-phosphate	$C_5H_9O_8P$	−2
r5p	α-D-Ribose 5-phosphate	$C_5H_9O_8P$	−2
f6p	D-Fructose 6-phosphate	$C_6H_{11}O_9P$	−2
g3p	Glyceraldehyde 3-phosphate	$C_3H_5O_6P$	−2
xu5p-D	D-Xylulose 5-phosphate	$C_5H_9O_8P$	−2
s7p	Sedoheptulose 7-phosphate	$C_7H_{13}O_{10}P$	−2
e4p	D-Erythrose 4-phosphate	$C_4H_7O_7P$	−2
nadp	Nicotinamide adenine dinucleotide phosphate ($NADP^+$)	$C_{21}H_{25}N_7O_{17}P_3$	−3
nadph	Nicotinamide adenine dinucleotide phosphate-reduced	$C_{21}H_{26}N_7O_{17}P_3$	−4

Under anaerobic conditions there is a greater flux through the nonoxidative pathway than through the oxidative pathway (93, 94).

The decarboxylating oxidative pathway is effectively irreversible. It consumes D-glucose 6-phosphate and after three reactions catalyzed by glucose-6-phosphate dehydrogenase, *G6PDH2r* (95), 6-phosphogluconolactonase, *PGL* (96), and phosphogluconate dehydrogenase, *GND* (97), produces D-ribulose 5-phosphate. The first and third reactions in this pathway each reduces one nicotinamide adenine dinucleotide phosphate ($NADP^+$) to NADPH. D-Ribulose 5-phosphate is then reversibly structurally rearranged into α-D-ribose 5-phosphate in the reaction ribose-5-phosphate isomerase, *RPI* (98). The oxidative branch of the pentose phosphate shunt is important for production of reducing power in the form of NADPH. However, the pentose phosphate shunt is not the only source of NADPH (99). The reactions catalyzed by NAD(P) transhydrogenase, *THD2*, isocitrate dehydrogenase (NADP), *ICDHyr*, and malic enzyme (NADP), *ME2*, can also supply *E. coli* with NADPH.

The nonoxidative reversible rearrangement of the glycolytic sugar monophosphates to the pentose phosphate shunt sugar monophosphates is a simple mechanism for creating these precursors (100), but it does not contribute to reducing power. This rearrangement requires three steps, as illustrated in Fig. 6. First, transketolase, *TKT2* (101, 102), catalyzes the conversion of a 6-carbon compound, D-fructose 6-phosphate, plus a 3-carbon compound, glyceraldehyde 3-phosphate, into a 5-carbon compound, D-xylulose 5-phosphate, plus a 4-carbon precursor, D-erythrose 4-phosphate. Then transaldolase, *TALA*, catalyzes the conversion of this 4-carbon compound plus another 6-carbon D-fructose 6-phosphate, into a 7-carbon, sedoheptulose 7-phosphate, plus another molecule of 3-carbon glyceraldehyde 3-phosphate. Next, the multifunctional enzyme transketolase, *TKT1*, catalyzes the conversion of this 7-carbon plus this 3-carbon into two different 5-carbon compounds, D-xylulose 5-phosphate and the precursor α-D-ribose 5-phosphate. In addition to ribose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, *RPE* (103), provides another reversible catalytic link between the oxidative and nonoxidative branches since it interconverts D-xylulose 5-phosphate and D-ribulose 5-phosphate.

Tricarboxylic Acid Cycle

The tricarboxylic acid (TCA) cycle is a well-studied pathway with a variety of functions depending on the environment (Table 5, Table 6, and Fig. 7). During aerobic growth on 6-carbon sugars such as glucose, the TCA cycle functions to create the precursors oxaloacetate, 2-oxoglutarate (also commonly called α-ketoglutarate), and succinyl-CoA. The aerobic production of biomass precursors is carried out primarily by the oxidative arm of the TCA cycle, from oxaloacetate to 2-oxoglutarate. This is the counterclockwise, lower part of the cycle in Fig. 7. The full TCA cycle, continuing counterclockwise in Fig. 7, can totally oxidize acetyl-CoA. Cycle intermediates are still required as biosynthetic precursors, so flux from anaplerotic pathways is required to maintain the pool of dicarboxylic intermediates ("Glyoxylate Cycle, Gluconeogenesis, and Anaplerotic Reactions," below). Under anaerobic conditions, the TCA cycle functions not as a cycle but as two separate pathways. The oxidative pathway, the counterclockwise lower part of the cycle in Fig. 7, forms the precursor 2-oxoglutarate. The reductive pathway, the clockwise upper part of the cycle in Fig. 7, forms the precursor succinyl-CoA. *E. coli* can grow in an environment where the only carbon substrate is one of the TCA cycle intermediates. This is enabled by proton symport transport reactions that translocate either 2-oxoglutarate, succinate, fumarate, or L-malate into the cell.

Pyruvate dehydrogenase, *PDH* (see Fig. 10), catalyzes the synthesis of acetyl-CoA from pyruvate and coenzyme A

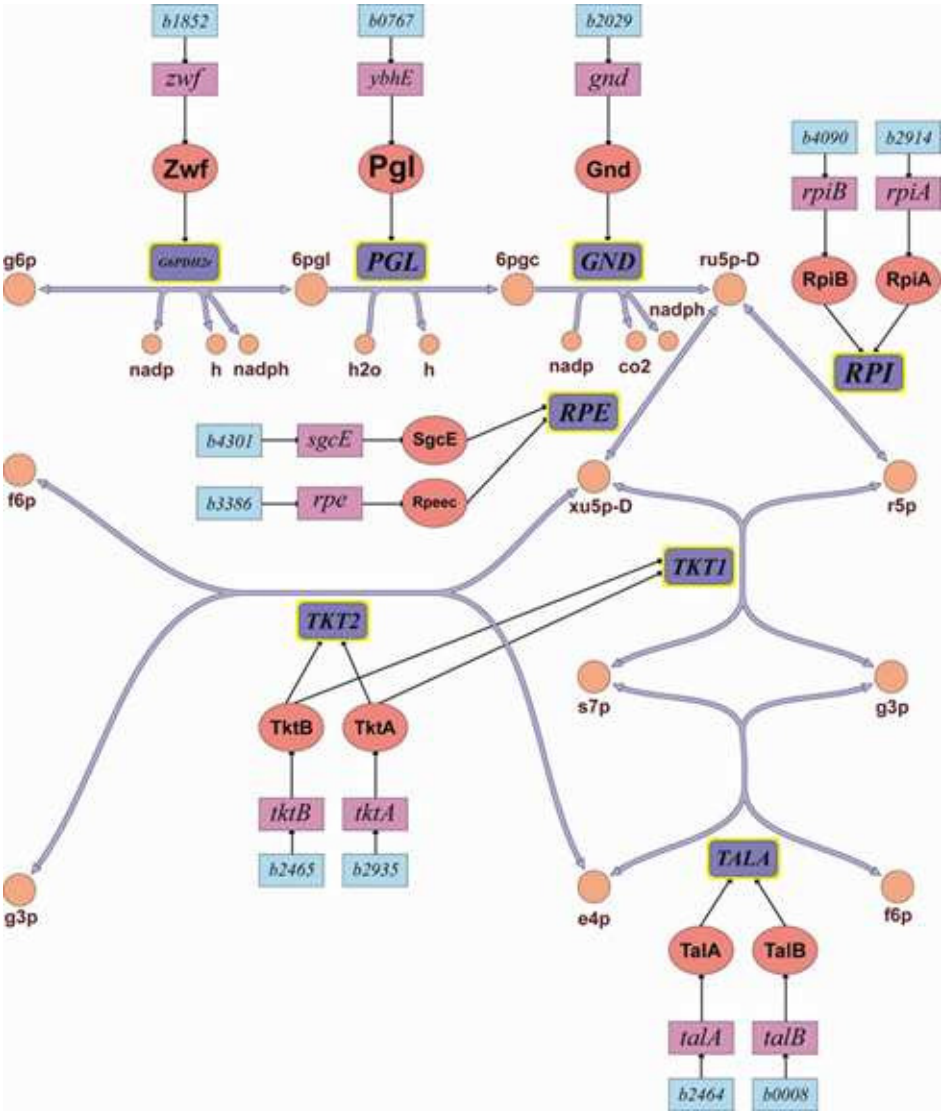

Figure 6 The pentose phosphate shunt. The irreversible, decarboxylating, oxidative pathway runs along the top, and below is the reversible set of rearrangements that constitute the nonoxidative pathway.

with concomitant reduction of NAD^+ to NADH. This reaction is carried out by a large multienzyme complex containing 12 AceE dimers, 6 Lpd dimers, and a 24-AceF core (104). Citrate synthase, *CS* (105), catalyzes the condensation reaction of the 2-carbon acetate residue from acetyl-CoA and a molecule of 4-carbon oxaloacetate to form the 6-carbon compound citrate, with the release of coenzyme A. Next, the reactions aconitase A, *ACONTa*, and aconitase B, *ACONTb*, isomerize citrate to isocitrate via the metabolite *cis*-aconitate (106, 107). Isocitrate sits at a branch point in the TCA cycle, where carbon flux either continues in the oxidative branch to the reaction catalyzed by isocitrate dehydrogenase, *ICDHyr* (108), or is diverted into the glyoxylate cycle by isocitrate lyase, *ICL*. The glyoxylate cycle is discussed further in "Glyoxylate Cycle, Gluconeogenesis, and Anaplerotic Reactions," below.

Isocitrate dehydrogenase catalyzes the decarboxylation of isocitrate, producing 2-oxoglutarate and CO_2 while reducing $NADP^+$ to NADPH. 2-Oxoglutarate provides a carbon backbone for synthesis of glutamate and glutamine, the central metabolites in nitrogen metabolism, discussed in "Nitrogen Metabolism." 2-Oxoglutarate dehydrogenase, *AKGDH*, catalyzes the decarboxylation of 2-oxoglutarate, producing CO_2, reducing NAD^+ to NADH, and transferring coenzyme A to the decarboxylated carbon compound to form succinyl-CoA.

Table 5 Tricarboxylic acid cycle reactions

Abbr.	Reaction	Equation
CS	Citrate synthase	accoa + h2o + oaa → cit + coa + h
ACONTa	Aconitase (half-reaction A, citrate hydro-lyase)	cit ⟷ acon-C + h2o
ACONTb	Aconitase (half-reaction B, isocitrate hydro-lyase)	acon-C + h2o ⟷ icit
ICDHyr	Isocitrate dehydrogenase (NADP)	icit + nadp ⟷ akg + co2 + nadph
AKGDH	2-Oxoglutarate dehydrogenase	akg + coa + nad → co2 + nadh + succoa
SUCOAS	Succinyl-CoA synthetase (ADP-forming)	atp + coa + succ → adp + pi + succoa
FRD7	Fumarate reductase	fum + q8h2 → q8 + succ
SUCDi	Succinate dehydrogenase (irreversible)	q8 + succ → fum + q8h2
FUM	Fumarase	fum + h2o ⟷ mal-L
MDH	Malate dehydrogenase	mal-L + nad ⟷ h + nadh + oaa
AKGt2r	2-Oxoglutarate reversible transport via symport	akg[e] + h[e] ⟷ akg + h
SUCCt3	Succinate transport out via proton antiport	h[e] + succ → h + succ[e]
SUCCt2_2	Succinate transport via proton symport (2 H)	2 h[e] + succ[e] → 2 h + succ
FUMt2_2	Fumarate transport via proton symport (2 H)	fum[e] + 2 h[e] → fum + 2 h
MALt2_2	Malate transport via proton symport (2 H)	2 h[e] + mal-L[e] → 2 h + mal-L

2-Oxoglutarate dehydrogenase is a large enzyme complex of the same family as pyruvate dehydrogenase, containing 12 SucA enzymes, an Lpd dimer, and 24 SucB enzymes (109).

During aerobic growth, the TCA cycle continues counterclockwise from succinyl-CoA in Fig. 7. Succinyl-CoA synthetase, *SUCOAS*, generates ATP by separating succinate and coenzyme A (110). Succinate dehydrogenase, *SUCDi*, is a multiprotein enzyme complex that straddles the cytoplasmic membrane allowing it to couple the TCA cycle to the electron transport chain (111), discussed in "Electron Transport Chain, Oxidative Phosphorylation, and Transfer of Reducing Equivalents." The succinate dehydrogenase complex catalyzes the irreversible oxidation of succinate to fumarate while reducing ubiquinone-8, q8, to ubiquinol-8, q8h2 (112). Ubiquinol-8 is then released from the enzyme complex and free to diffuse through the cytoplasmic membrane to interact with subsequent enzymes of the electron transport chain. After succinate is converted fumarate, fumarase, *FUM*, reversibly catalyzes the conversion of fumarate and water into L-malate (113, 114, 115). Finally, to complete the TCA cycle, malate dehydrogenase, *MDH*, reversibly catalyzes the conversion of malate into oxaloacetate while reducing NAD^+ to NADH (116). When this set of reactions is used in reverse, clockwise in Fig. 7, as the reductive pathway of the TCA cycle, a reversing reaction is catalyzed by fumarate reductase, *FRD7* (117). In the model, this reaction oxidizes ubiquinol-8 to ubiquinone-8, although in actual *E. coli*, fumarate reductase oxidizes the electron carrier menaquinol-8 instead (118). This reaction had to be included in an unrealistic form because the simplified electron transport system in the model includes only ubiquinone-8/ubiquinol-8 as an electron carrier.

Glyoxylate Cycle, Gluconeogenesis, and Anaplerotic Reactions

When growing on some substrates, the glyoxylate cycle is used instead of the TCA cycle because it bypasses the reactions that lose carbon in the form of CO_2. The glyoxylate cycle consists of some of the reactions in the TCA cycle as well as other reactions used only by the glyoxylate cycle (see Table 7, Table 8, and Fig. 8). It overlaps with the TCA cycle from the incorporation of acetyl-CoA to the production of isocitrate at the aforementioned branch point in the oxidative arm of the TCA cycle. Isocitrate lyase, *ICL*, catalyzes the cleavage of 6-carbon isocitrate into 4-carbon succinate and 2-carbon glyoxylate (119). Malate synthase, *MALS*, then catalyzes the condensation of glyoxylate with another acetyl-CoA,

Table 6 Tricarboxylic acid cycle metabolites

Abbr.	Metabolite	Formula	Charge
accoa	Acetyl-CoA	$C_{23}H_{34}N_7O_{17}P_3S$	−4
cit	Citrate	$C_6H_5O_7$	−3
acon-C	*cis*-aconitate	$C_6H_3O_6$	−3
icit	Isocitrate	$C_6H_5O_7$	−3
akg	2-Oxoglutarate	$C_5H_4O_5$	−2
succoa	Succinyl-CoA	$C_{25}H_{35}N_7O_{19}P_3S$	−5
fum	Fumarate	$C_4H_2O_4$	−2
mal-L	L-Malate	$C_4H_4O_5$	−2
oaa	Oxaloacetate	$C_4H_2O_5$	−2
coa	Coenzyme A	$C_{21}H_{32}N_7O_{16}P_3S$	−4
q8	Ubiquinone-8	$C_{49}H_{74}O_4$	0
q8h2	Ubiquinol-8	$C_{49}H_{76}O_4$	0

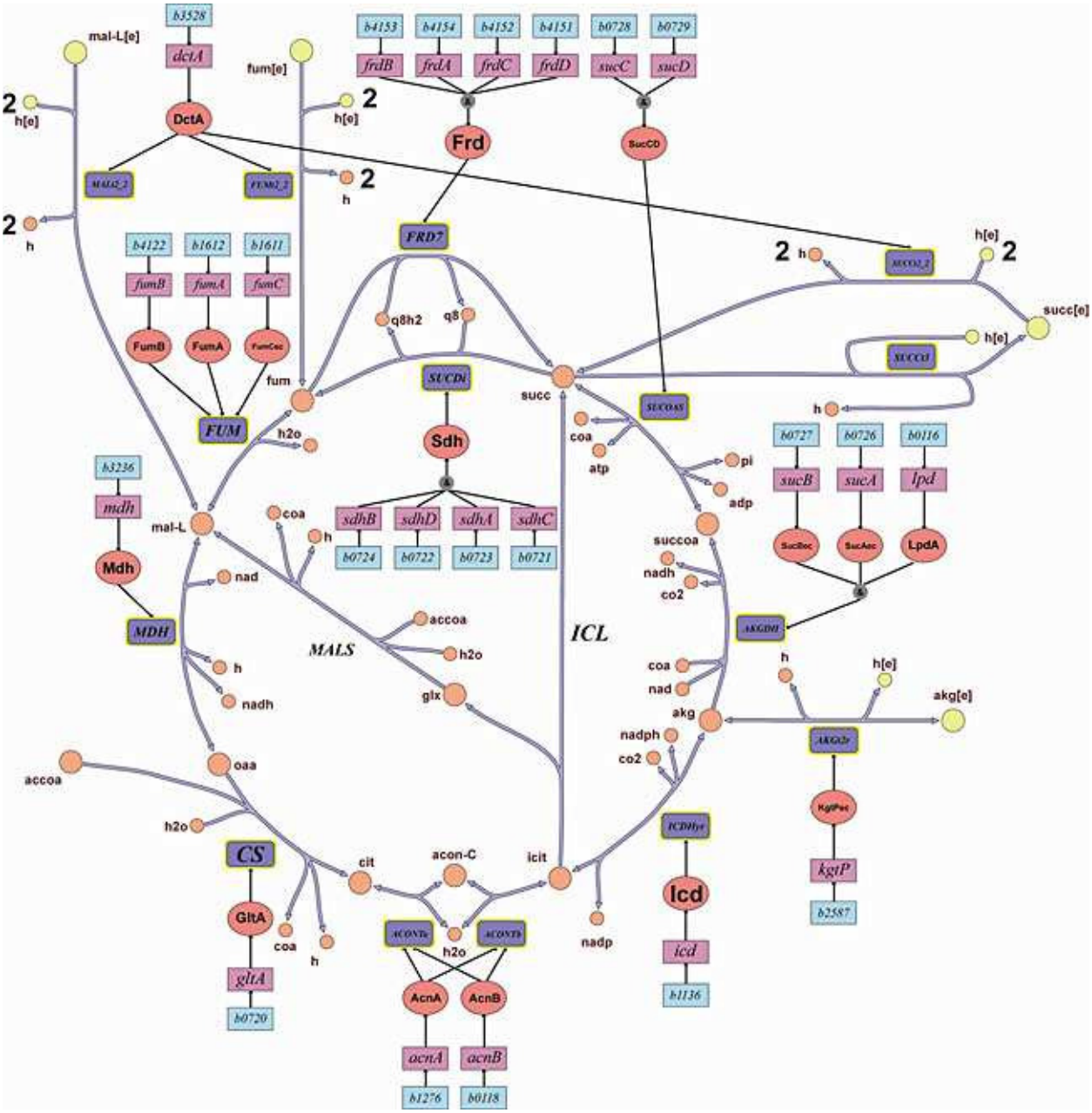

Figure 7 The TCA cycle reactions. The oxidative pathway of the TCA cycle runs counterclockwise in the lower part of the cycle, from oxaloacetate (oaa) through 2-oxoglutarate (akg). Continuing counterclockwise from 2-oxoglutarate, the full TCA cycle can totally oxidize acetyl-CoA, or it can function as two separate oxidative and reductive pathways. The reactions in the center of the cycle, *ICL* and *MALS*, are part of the glyoxylate cycle, discussed in Glyoxylate Cycle, Gluconeogenesis, and Anaplerotic Reactions.

yielding malate (120). Succinate generated in the first step can continue along the TCA cycle to eventually form oxaloacetate.

In addition to growing on hexoses or pentoses, *E. coli* can also grow on 2-, 3-, or 4-carbon sources, such as lactate or malate, but in this situation, some 6-carbon metabolites in the glycolytic pathway are still required as precursors for biomass components. With only 2-, 3-, or 4-carbon sources in the environment, the glycolytic pathway can actually be reversed to produce net flux from pyruvate to glucose 6-phosphate. This reversal of glycolytic flux is referred to as gluconeogenesis. The two reactions of glycolysis that are effectively irreversible, catalyzed by phosphofructokinase, *PFK*, and pyruvate kinase, *PYK*, are replaced with two reversing reactions, catalyzed by

Table 7 Glyoxylate cycle, anaplerotic reactions, and gluconeogenesis reactions

Abbr.	Reaction	Equation
ICL	Isocitrate lyase	icit → glx + succ
MALS	Malate synthase	accoa + glx + h2o → coa + h + mal-L
ME1	Malic enzyme (NAD)	mal-L + nad → co2 + nadh + pyr
ME2	Malic enzyme (NADP)	mal-L + nadp → co2 + nadph + pyr
PPS	Phosphoenolpyruvate synthase	atp + h2o + pyr → amp + 2 h + pep + pi
PPCK	Phosphoenolpyruvate carboxykinase	atp + oaa → adp + co2 + pep
PPC	Phosphoenolpyruvate carboxylase	co2 + h2o + pep → h + oaa + pi

fructose-bisphosphatase, *FBP*, and phosphoenolpyruvate synthase, *PPS*, respectively. Phosphoenolpyruvate synthase (121) catalyzes the conversion of pyruvate to phosphoenolpyruvate and in the process hydrolyzes one ATP to AMP (122).

Anaplerotic reactions replenish TCA cycle intermediates drained off for biosynthesis. The TCA cycle operating cyclically can completely oxidize acetate to carbon dioxide without net consumption or production of intermediates. However, intermediates of the TCA cycle such as oxaloacetate and 2-oxoglutarate are consumed in the production of macromolecules. TCA cycle intermediate generation from the glycolytic metabolites is accomplished by the irreversible carbon dioxide-fixing conversion of the 3-carbon phosphoenolpyruvate to the 4-carbon oxaloacetate, catalyzed by the enzyme phosphoenolpyruvate carboxylase, *PPC* (Fig. 8).

Growth on 4-carbon dicarboxylic acid intermediates of the TCA cycle, such as malate, requires that the cell be able to produce phosphoenolpyruvate for gluconeogenesis. Two pathways exist to fulfill these phosphoenolpyruvate demands (123, 124). One pathway involves the conversion of malate to pyruvate by malic enzyme, *ME1* or *ME2* (125, 126), followed by the synthesis of phosphoenolpyruvate from pyruvate by phosphoenolpyruvate synthase, *PPS* (127). Malic enzyme, *ME1*, reduces one molecule of NAD^+ to NADH while converting malate to pyruvate. A second parallel reaction, malic enzyme (NADP), *ME2*, reduces one molecule of $NADP^+$ to NADPH. The other pathway from the TCA cycle to glycolytic intermediates is the conversion of oxaloacetate to phosphoenolpyruvate by the action of phosphoenolpyruvate carboxykinase, *PPCK* (128). Phosphoenolpyruvate carboxykinase catalyzes the reverse reaction to the anaplerotic enzyme, phosphoenolpyruvate carboxylase, *PPC* (129). The former reaction consumes a high-energy phosphate bond in ATP and produces CO_2, whereas phosphoenolpyruvate carboxylase releases inorganic phosphate and consumes CO_2. Although the reactions catalyzed by the enzymes phosphoenolpyruvate carboxykinase and malic enzyme are thermodynamically reversible, physiologically they are found to operate unidirectionally (130, 131).

Electron Transport Chain, Oxidative Phosphorylation, and Transfer of Reducing Equivalents

The electron transport chain and oxidative phosphorylation are used to produce the bulk of the cell's ATP under aerobic conditions. The electron transport chain translocates protons (H^+) from the cytoplasm across the cytoplasmic membrane into the periplasmic space (see Table 9, Table 10, and Fig. 9). Since the cytoplasmic membrane is effectively impermeable to protons and hydroxyl ions (OH^-), this establishes a difference in concentration of protons, and a difference in electrical charge, across the cytoplasmic membrane. This thermodynamic potential difference gives rise to a proton motive force that can be utilized to drive a myriad of endergonic reactions, such as synthesis of high-energy currency metabolites like ATP. In the model, protons are translocated into the extracellular medium as a simplification, but this is a reasonable approximation given that the pH of the periplasm and that of extracellular medium are the same (132).

The electron transport chain of *E. coli* consists of several different respiratory dehydrogenases, quinones, and terminal reductases. There are 15 different dehydrogenases that accept electrons from donors such as NADH or succinate, and then pass the electrons to one of three different quinones, which then deliver the electrons to one of at least 14 different terminal reductases. The reductases complete the chain by reducing a terminal electron acceptor such as oxygen or fumarate. Some, but not all dehydrogenases and reductases pump protons

Table 8 Glyoxylate metabolite

Abbr.	Metabolite	Formula	Charge
glx	Glyoxylate	C_2HO_3	−1

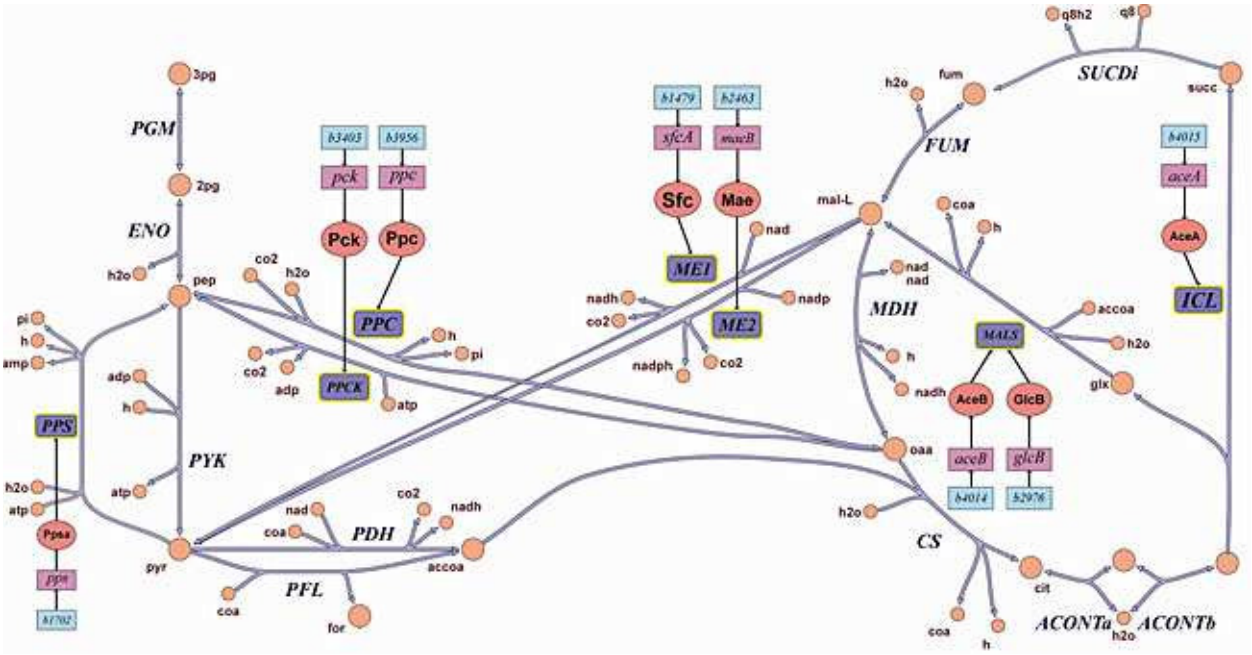

Figure 8 Glyoxylate cycle, gluconeogenic, and anaplerotic reactions. Acetate enters the glyoxylate cycle via acetyl-CoA, accoa, at two points: lower center left entering the citrate synthase reaction, *CS*, and in the center of the TCA cycle, entering the malate synthase reaction, *MALS*. Tricarboxylic acid intermediates are then converted into gluconeogenic precursors via either malic enzyme, *ME1* and *ME2*, or phosphoenolpyruvate carboxykinase, *PPCK*. Gluconeogenesis then synthesizes 6-carbon sugars with the aid of reactions such as phosphoenolpyruvate synthase, *PPS*, which effectively reverses an otherwise thermodynamically unfavorable glycolytic reaction. Phosphoenolpyruvate carboxylase, *PPC*, is an anaplerotic reaction that uses glycolytic intermediates to replenish the precursors drained from the TCA cycle for biosynthesis.

into the periplasm. It is possible for the dehydrogenases, quinones, and reductases to be used in many different combinations, so the entire system can be very complicated (133). The core *E. coli* model condenses the sequence of steps in the electron transport chain into two reactions, representing generic NADH dehydrogenase and cytochrome oxidase reactions, connected by only one quinone. First, an NADH dehydrogenase, *NADH16*, catalyzes the oxidation of NADH to form NAD^+ while removing four protons from the cytoplasm, translocating three protons to the extracellular space and combining the fourth with a proton plus two electrons from NADH with ubiquinone-8 to form the reduced ubiquinol-8 (134, 135). Ubiquinone-8 and ubiquinol-8 are oil-soluble coenzymes that diffuse freely within the lipid environment of the cytoplasmic membrane (Fig. 9). The next condensed step is when cytochrome oxidase, *CYTBD*, catalyzes the oxidation of ubiquinol-8 back to ubiquinone-8, which drives the translocation of two more protons into the extracellular space (136). The two spare electrons are then combined with two cytoplasmic protons and an oxygen atom to form water. Oxygen spontaneously diffuses from the environment into the cell down a concentration gradient, *O2t*, and represents an exogenous source of terminal electron acceptor.

Table 9 Electron transport chain, oxidative phosphorylation, and transfer of reducing equivalents reactions

Abbr.	Reaction	Equation
NADH16	NADH dehydrogenase (ubiquinone-8 & 3 protons)	4 h + nadh + q8 → 3 h[e] + nad + q8h2
CYTBD	Cytochrome oxidase bd (ubiquinol-8: 2 protons)	2 h + ½ o2 + q8h2 → h2o + 2 h[e] + q8
O2t	o2 transport via diffusion	o2[e] ⟷ o2
ATPS4r	ATP synthase (four protons for one ATP)	adp + 4 h[e] + pi ⟷ atp + h2o + 3 h
ATPM	ATP maintenance requirement	atp + h2o → adp + h + pi
ADK1	Adenylate kinase	amp + atp ⟷ 2 adp
THD2	NAD(P) transhydrogenase	2 h[e] + nadh + nadp → 2 h + nad + nadph
NADTRHD	NAD transhydrogenase	nad + nadph → nadh + nadp

The enzyme ATP synthase, *ATPS4r*, catalyzes the synthesis of ATP from ADP, forming a high-energy phosphate bond by coupling catalysis to the import of four protons that were pumped out by the electron transport

Table 10 Electron transport chain, oxidative phosphorylation, and transfer of reducing equivalents metabolites

Abbr.	Metabolite	Formula	Charge
q8	Ubiquinone-8	$C_{49}H_{74}O_4$	0
q8h2	Ubiquinol-8	$C_{49}H_{76}O_4$	0
nad	Nicotinamide adenine dinucleotide (NAD^+)	$C_{21}H_{26}N_7O_{14}P_2$	−1
nadh	Nicotinamide adenine dinucleotide-reduced	$C_{21}H_{27}N_7O_{14}P_2$	−2
nadp	Nicotinamide adenine dinucleotide phosphate ($NADP^+$)	$C_{21}H_{25}N_7O_{17}P_3$	−3
nadph	Nicotinamide adenine dinucleotide phosphate-reduced	$C_{21}H_{26}N_7O_{17}P_3$	−4
atp	Adenosine triphosphate	$C_{10}H_{12}N_5O_{13}P_3$	−4
adp	Adenosine diphosphate	$C_{10}H_{12}N_5O_{10}P_2$	−3
amp	Adenosine monophosphate	$C_{10}H_{12}N_5O_7P$	−2
h	H^+	H	1

chain (137, 138). The exact number of high-energy phosphate bonds that are generated per oxygen atom used as a terminal acceptor is the P/O ratio. This varies depending on periplasmic pH and other environmental conditions, but the core model P/O ratio is stoichiometrically fixed at 1.25. The ATP maintenance reaction, *ATPM*, is not a real biochemical reaction. It is included for modeling purposes since the scope of the *E. coli* core model does not extend to all of the reactions that consume ATP in the cell. Adenylate kinase, *ADK1*, is a phosphotransferase enzyme that catalyzes the interconversion of adenine nucleotides, and plays an important role in cellular energy homeostasis (139, 140).

NADH is used for the catabolic activities of the cell, for instance, driving the export of protons into the periplasmic space in the electron transport chain in the reaction catalyzed by NADH dehydrogenase, *NADH16*. In contrast, NADPH is essential for anabolic metabolism such as the biosynthesis of building blocks for polymerization reactions from precursor metabolites produced by the fueling pathways of core metabolism. Maintaining the proper balance between anabolic reduction charge, NADPH/$NADP^+$, and catabolic reduction charge, NADH/NAD^+, is achieved by reactions catalyzed by transhydrogenase enzymes. NAD(P) transhydrogenase, *THD2*, catalyzes the transfer of a hydride ion from NADH to create NADPH, in a reaction coupled to the proton motive force (141). The opposite transfer, of a hydride ion from NADPH to create NADH, is catalyzed by another enzyme, NAD transhydrogenase, *NADTRHD*,

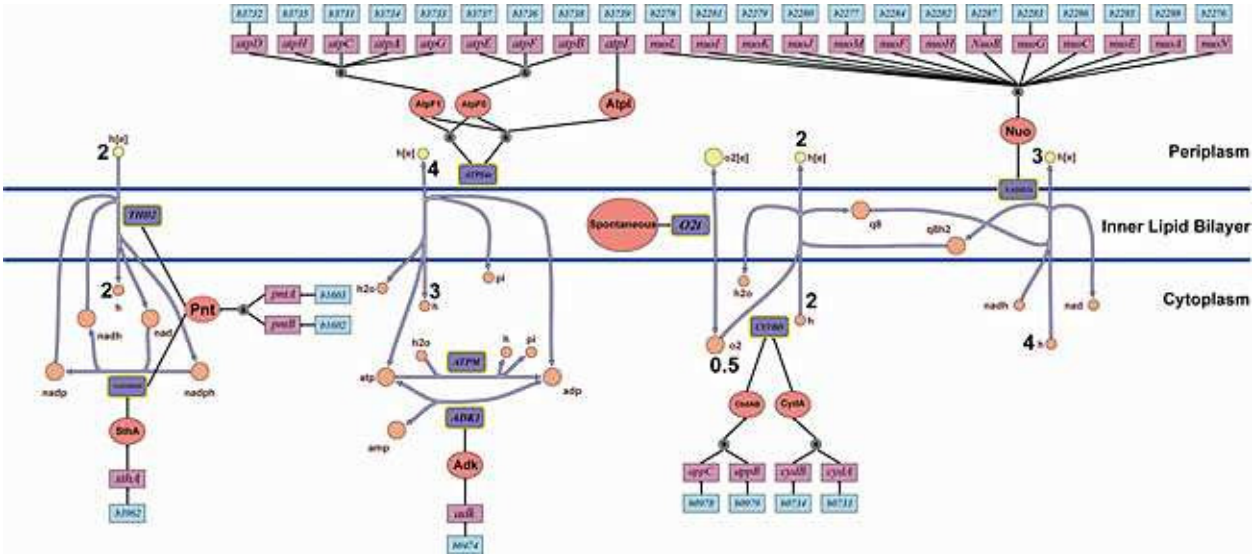

Figure 9 NAD(P) transhydrogenase, *THD2*, and NAD transhydrogenase, *NADTRHD*, catalyze the transfer of reducing equivalents between anabolic and catabolic reduction charge (far left). The electron transport chain (right) generates an electrochemical potential between the cytoplasm and periplasm. Oil-soluble ubiquinone-8 (q8) and ubiquinol-8 (q8h2) freely diffuse within the lipid membrane separating the periplasm and cytoplasm. Enzyme complexes that involve proton translocation straddle the cytoplasmic membrane and are in contact with periplasm and cytoplasm. ATP synthase, *ATPS4r*, catalyzes the synthesis of ATP from ADP by using the electrochemical potential between the cytoplasm and the periplasm (middle).

but it is not coupled to the translocation of protons (142). This pair of reactions effectively allows the transfer of reducing equivalents between anabolic and catabolic reduction charge.

Fermentation

During aerobic respiration, oxygen is the terminal electron acceptor for the electron transport chain, and the ATP required for biosynthesis is produced by ATP synthase. Under anaerobic conditions, *E. coli* can generate ATP by substrate-level phosphorylation in the process of fermentation, where excess carbon is secreted as various organic by-products (see Table 11, Table 12, and Fig. 10). Glycolysis results in the net production of 2 ATP per glucose by substrate-level phosphorylation, but this is very low compared with the 17.5 ATP per glucose generated during aerobic respiration (in the model 17.5 ATP per glucose is generated, but this number can vary in vivo). The substrates of fermentation are typically sugars, so during fermentative growth, each cell must maintain a large flux through glycolysis to generate sufficient ATP to drive the constitutive biosynthesis, polymerization, and assembly reactions required for growth. This necessitates a large efflux of fermentative end products since there is insufficient ATP to assimilate all carbon as biomass. Approximately 10% of carbon substrate is assimilated because of the poor energy yield of fermentation (143). Glycolysis also produces two molecules of NADH for each glucose; therefore, NADH must be reoxidized by fermentation to regenerate NAD^+ to maintain the oxidation-reduction balance of the cell.

Table 11 Fermentation reactions

Abbr.	Reaction	Equation
LDH_D	D-Lactate dehydrogenase	lac-D + nad ⟷ h + nadh + pyr
D_LACt2	D-Lactate transport via proton symport	h[e] + lac-D[e] ⟷ h + lac-D
PDH	Pyruvate dehydrogenase	coa + nad + pyr → accoa + co2 + nadh
PFL	Pyruvate formate lyase	coa + pyr → accoa + for
FORti	Formate transport via diffusion	for → for[e]
FORt2	Formate transport via proton symport	for[e] + h[e] → for + h
PTAr	Phosphotransacetylase	accoa + pi ⟷ actp + coa
ACKr	Acetate kinase	ac + atp ⟷ actp + adp
ACALD	Acetaldehyde dehydrogenase (acetylating)	acald + coa + nad ⟷ accoa + h + nadh
ALCD2x	Alcohol dehydrogenase (ethanol)	etoh + nad ⟷ acald + h + nadh
ACt2r	Acetate reversible transport via proton symport	ac[e] + h[e] ⟷ ac + h
ACALDt	Acetaldehyde reversible transport	acald[e] ⟷ acald
ETOHt2r	Ethanol reversible transport via proton symport	etoh[e] + h[e] ⟷ etoh + h

Table 12 Fermentation metabolites

Abbr.	Metabolite	Formula	Charge
lac-D	D-Lactate	$C_3H_5O_3$	−1
for	Formate	CHO_2	−1
actp	Acetyl-phosphate	$C_2H_3O_5P$	−2
ac	Acetate	$C_2H_3O_2$	−1
acald	Acetaldehyde	C_2H_4O	0
etoh	Ethanol	C_2H_6O	0

E. coli fermentation normally generates a mixture of end products anaerobically from sugars. The major soluble products are acetate, ethanol, lactate, and formate, with a smaller amount of succinate (143). In addition, fermentation results in the production of substantial quantities of carbon dioxide and hydrogen. Depending on the pH of the culture medium, and the redox state of the fermentation substrate, a cell may vary the relative flux through each of the fermentation pathways branching from pyruvate. When the pH of the environment drops because of increased concentrations of other acidic fermentative end products, such as acetic, formic, or succinic acid, then flux may be increased through the reaction catalyzed by D-lactate dehydrogenase, *LDH_D* (144, 145). This results in the reduction of pyruvate to form lactate while oxidizing NADH to form NAD^+.

Pyruvate formate lyase, *PFL*, catalyzes the nonoxidative cleavage of pyruvate into acetyl-CoA and formate, with the incorporation of coenzyme A into acetyl-CoA (146, 147). Acetyl-CoA can then lead to either of the fermentative end products, acetate or ethanol. Both pathways involve two-step reversible mechanisms. In the conversion to acetate, phosphotransacetylase, *PTAr* (148), catalyzes transfer of a phosphate group onto the acetyl moiety of acetyl-CoA, to form acetyl-phosphate and release coenzyme A to be recycled for the previous

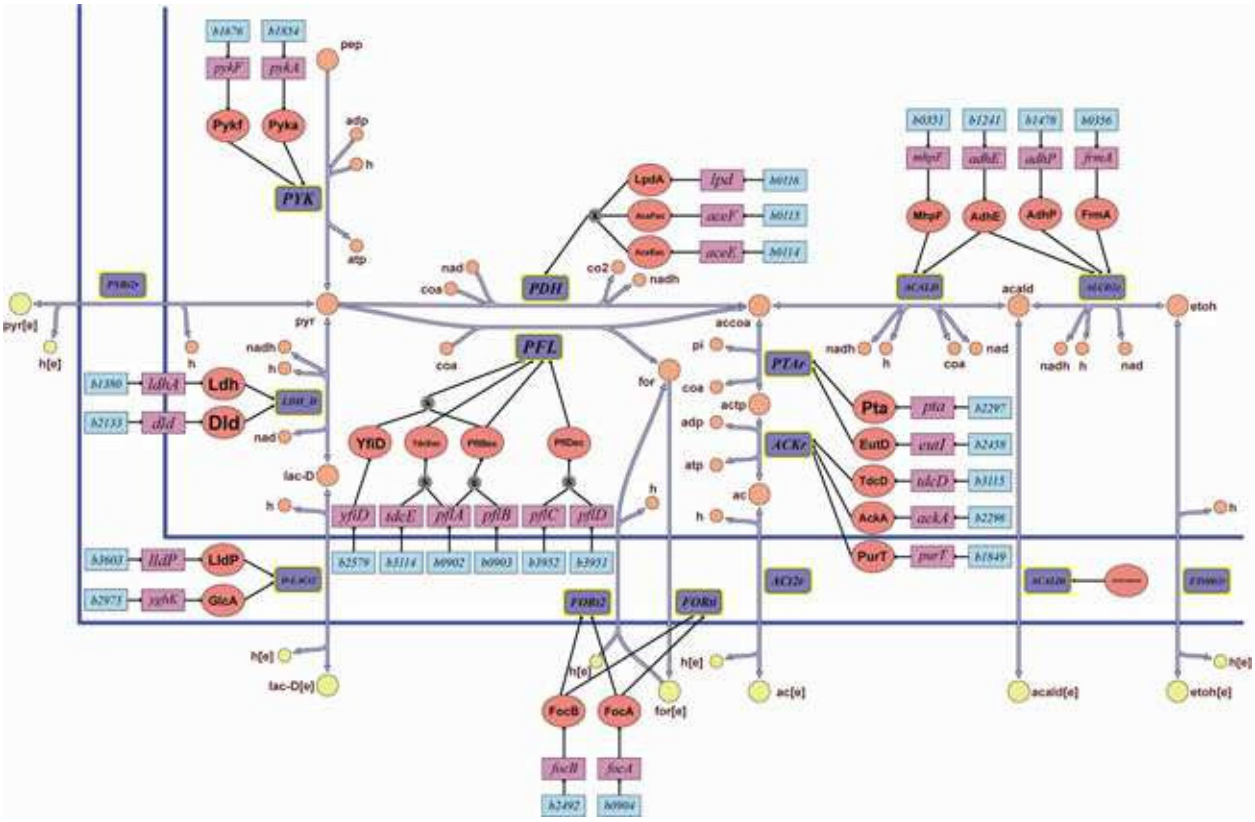

Figure 10 Fermentation begins with pyruvate. It can be converted to D-lactate (lac-D) and excreted, or it can be converted to formate (for), acetate (ac), acetaldehyde (acald), or ethanol (etoh).

reaction catalyzed by pyruvate formate lyase, *PFL*. Then, acetate kinase, *ACKr*, catalyzes the conversion of acetyl-phosphate to acetate, in the process forming a much needed high-energy phosphate bond by converting ADP to ATP (149). The reactions catalyzed by phospho-transacetylase and acetate kinase are thermodynamically reversible. This allows *E. coli* to grow aerobically on acetate by reversing the flux through this fermentative pathway. During the conversion of acetyl-CoA to ethanol, two molecules of NADH are reoxidized, one by the first reaction catalyzed by acetaldehyde dehydrogenase (acetylating), *ACALD* (150), and the second by the subsequent reaction catalyzed by alcohol dehydrogenase (ethanol), *ALCD2x* (151). Fermentation of many different substrates is possible because ethanol is more reduced than sugars, whereas acetate is more oxidized than ethanol. Redox balancing is achieved by varying the ratio of ethanol to acetate secreted. The end products of each fermentation pathway exit the cell along a concentration gradient, and, in the process, transport a proton from the cytoplasm to the periplasmic space.

Nitrogen Metabolism

Nitrogen is the fourth most abundant element in *E. coli* and enters the cell either by ammonium ion uptake, *NH4t*, or as a moiety within organic molecules such as the amino acid L-glutamine or L-glutamate. Glutamate is an extremely abundant metabolite, with a measured concentration of 100.55 μmol per gram dry weight (gDW^{-1}) and production rate of 4.77 mmol gDW^{-1} h^{-1}. Glutamine is less abundant, with a measured concentration of 3.92 μmol gDW^{-1} and production rate of 3.36 mmol gDW^{-1} h^{-1} (152). Glutamate is a central component of the nitrogen metabolism of *E. coli*, as it is a nitrogen donor in many reactions. In particular, it is required for the synthesis of the other amino acids because its amino group is transferred to other compounds in transamination reactions. Glutamine is actively transported across the inner cytoplasmic membrane by an ATP-binding cassette transporter, *GLNabc*, which imports one molecule of glutamine while consuming one ATP (153). Glutamate is actively transported across the inner cytoplasmic membrane by proton symport, *GLUt2r* (154).

Direct ammonia assimilation is catalyzed by NADPH-specific reductive amination of 2-oxoglutarate to glutamate by glutamate dehydrogenase (NADP), *GLUDy* (152). Indirect ammonia assimilation is by a cyclic pair of sequential reactions catalyzed by glutamine synthetase, *GLNS* (155), and glutamate synthase (NADPH), *GLUSy* (152). Glutamine synthetase first catalyzes the assimilation of ammonia by converting glutamate, with one amino moiety, into glutamine, with two amino moieties. Then glutamate synthase, *GLUSy*, catalyzes the transfer of the amide group of glutamine to 2-oxoglutarate, generating a second molecule of glutamate (see Fig. 11, Table 13, and Table 14).

Mathematical Formulation of Core *E. coli* Metabolism

The core *E. coli* metabolic reconstruction can be converted into a mathematical model by encoding all of the metabolites and reactions in a stoichiometric matrix:

$$[\mathbf{S} \varepsilon Z^{m,n}]$$

The stoichiometric matrix for the core *E. coli* model has $m = 72$ rows, each corresponding to a metabolite, and $n = 95$ columns, each corresponding to a reaction. The coefficients within a single column of the stoichiometric matrix represent the stoichiometry of a single reaction. A

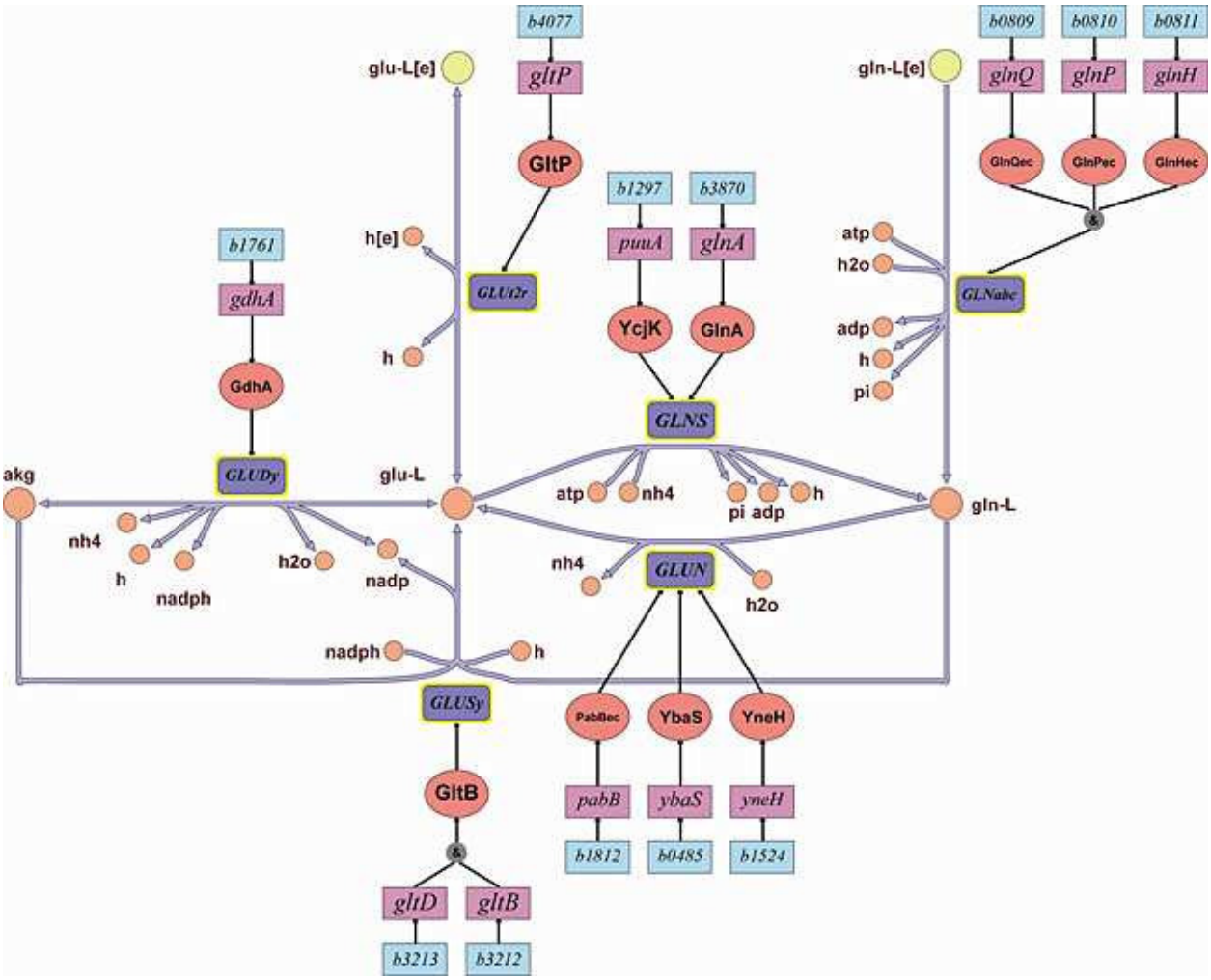

Figure 11 Nitrogen metabolism in *E. coli*. Glutamate dehydrogenase (NADP), *GLUDy*, catalyzes direct ammonia assimilation by NADPH-specific reductive amination of 2-oxoglutarate (akg) to L-glutamate (glu-L). Indirect ammonia assimilation is by a cyclic pair of sequential reactions, beginning with glutamine synthetase, *GLNS*, which catalyzes the addition of an amino moiety to glutamate to form glutamine (gln-L). Then, glutamate synthase (NADPH), *GLUSy*, catalyzes the transfer of this amino moiety to 2-oxoglutarate, generating two molecules of glutamate. The net result is amination of 2-oxoglutarate to glutamate, as in the direct pathway.

Table 13 Nitrogen metabolism reactions

Abbr.	Reaction	Equation
GLNabc	L-Glutamine transport via ABC system	atp + gln-L[e] + h2o → adp + gln-L + h + pi
GLUt2r	L-Glutamate transport via proton symport	glu-L[e] + h[e] ⟷ glu-L + h
GLUDy	Glutamate dehydrogenase (NADP)	glu-L + h2o + nadp ⟷ akg + h + nadph + nh4
GLNS	Glutamine synthetase	atp + glu-L + nh4 → adp + gln-L + h + pi
GLUSy	Glutamate synthase (NADPH)	akg + gln-L + h + nadph → 2 glu-L + nadp
GLUN	Glutaminase	gln-L + h2o → glu-L + nh4

negative stoichiometric coefficient indicates the number of molecules of a particular metabolite consumed in that reaction, and a positive stoichiometric coefficient represents the number of molecules of a metabolite produced in that reaction. Since each reaction typically involves only a few metabolites, the stoichiometric matrix is sparse, consisting mostly of zero coefficients. Each reaction must be balanced with respect to consumption and production of both chemical elements and electrical charge. The exceptions to this rule are the columns of the *E. coli* core stoichiometric matrix that correspond to exchange reactions at the boundary of the model and its environment. Each exchange reaction represents the influx or efflux of an extracellular metabolite between the model and the environment. A metabolite leaving the model is represented with a positive exchange flux, and the opposite for a metabolite entering the model. The biomass reaction, representing the synthesis of biomass, is also represented by a column of the stoichiometric matrix. The biomass reaction may be considered a special type of exchange reaction, and is discussed further in "Biomass Reaction."

The net flux through all of the 95 reactions in the model can be mathematically represented by a 95 dimensional vector:

$[\mathbf{v} \varepsilon R^n]$

By convention, net flux is represented with the unit millimoles per gram dry weight per hour (mmol gDW^{-1} h^{-1}). The dot product of the stoichiometric matrix and the net flux vector gives an m-dimensional vector of changes in metabolite concentration over time:

$[dx/dt \varepsilon R^m]$

During balanced growth, when considering a large population of cells, it is reasonable to assume that all metabolite concentrations are constant; therefore, we have the fundamental system of m equations for mass conservation at steady state:

$$\mathbf{S} \cdot \mathbf{v} = (dx/dt) = \mathbf{0} \quad (1)$$

Let the first row of the stoichiometric matrix correspond to ATP, then consider the meaning of the first equation in the system of equation 1:

$$\mathbf{S}_{1,1}\mathbf{v}_1 + \mathbf{S}_{1,2}\mathbf{v}_2 + \ldots + \mathbf{S}_{1,n}\mathbf{v}_n = (d[\mathbf{ATP}]/dt) = 0 \quad (2)$$

Since the stoichiometric coefficients represent the number of molecules consumed or produced in each reaction in the model, and the net fluxes represent the rates for each respective reaction, then equation 2 balances the total net production and consumption of *ATP* (Fig. 12).

The net flux for all biochemical reactions is bounded by diffusion limitations (156). In addition, experimental data on maximum reaction rates, substrate uptake rates, or waste secretion rates may be used to further tighten these bounds, mathematically represented by:

$$\mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub}$$

For reactions that are considered to be effectively thermodynamically irreversible in the forward direction, the convention is to set the lower bound on a reaction rate to zero (4). In any realistic stoichiometric model, there are more columns than rows in the stoichiometric matrix. Therefore, there is an insufficient number of equations to specify a unique net flux in equation 1. To predict a biologically meaningful net flux, flux balance analysis can be used (2, 49). Flux balance analysis uses linear programming to optimize a biologically motivated objective function, subject to steady-state mass balance and bounds on net fluxes to predict an in vivo flux. This can

Table 14 Nitrogen metabolites

Abbr.	Metabolite	Formula	Charge
glu-L	L-Glutamate	$C_5H_8NO_4$	−1
gln-L	L-Glutamine	$C_5H_{10}N_2O_3$	0
nh4	Ammonium	H_4N	1

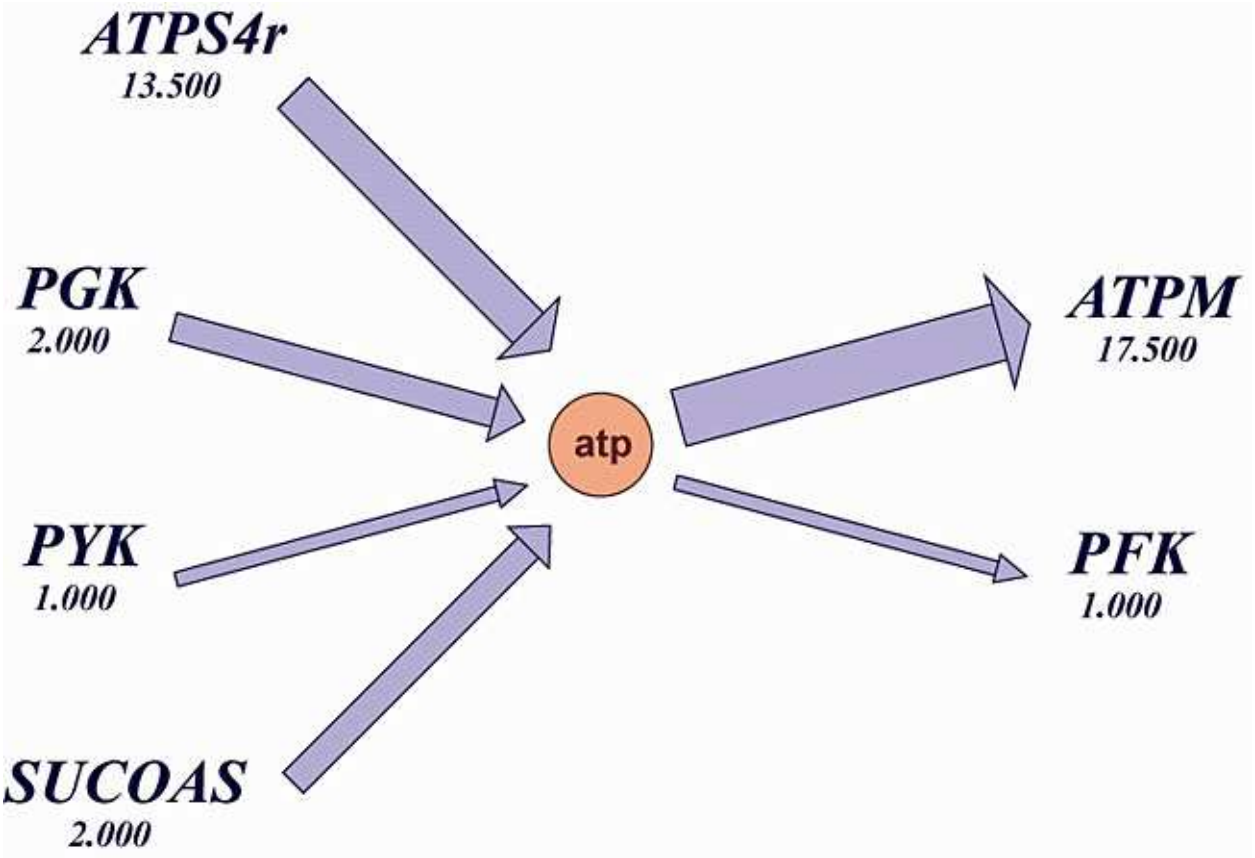

Figure 12 All of the reactions that produce or consume ATP in a particular flux distribution. The total amount of ATP produced is exactly equal to the amount of ATP consumed, so $d[\mathrm{ATP}]/dt = 0$.

be expressed as the following linear programming problem:

maximize: $\mathbf{c}^{\mathrm{T}} \cdot \mathbf{v}$

subject to: $\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$

$\mathbf{v}_{lb} \leq \mathbf{v} \leq \mathbf{v}_{ub}$

The objective function, $\mathbf{c}^{\mathrm{T}}\mathbf{v}$, can be any linear combination of reaction fluxes. In flux balance analysis of *E. coli*, the biomass reaction is often the sole reaction to be maximized by the objective function since maximum growth is evolutionarily favored in genetically heterogeneous culture (20). The publicly available COBRA Toolbox software for MATLAB (157) may be used to make numerical predictions of in vivo net flux by implementing a mathematical model as a computational model (158).

Biomass Reaction

To represent growth, the core *E. coli* model includes a biomass reaction that drains precursor metabolites from the network at stoichiometrically fixed relative rates while also producing several by-product metabolites (4) (see Table 15 and Fig. 13). These precursors are used to produce the lipids, proteins, nucleic acids, and other macromolecules required to replicate a cell. To determine these metabolites and their quantity, we used the dry weight composition data for an average *E. coli* B/r cell growing exponentially at 37°C under aerobic conditions in glucose minimal medium, with an approximate doubling time of 40 min having a dry cell weight of 2.81013 g (159). Since most of the subunits of the cellular macromolecules, such as nucleic acids and amino acids, are not present in the core model, they could not be directly accounted for in the biomass reaction. The metabolites in the core model that those macromolecular subunits are synthesized from are included instead. These

Table 15 Twenty-three different metabolites consumed or produced to simulate growth in the biomass reaction[a]

Abbr.	Metabolite	Stoichiometry
3pg	3-Phospho-D-glycerate	−1.496
accoa	Acetyl-CoA	−3.7478
adp	Adenosine diphosphate	59.81
akg	2-Oxoglutarate	4.1182
atp	Adenosine triphosphate	−59.81
coa	Coenzyme A	3.7478
e4p	D-Erythrose 4-phosphate	−0.361
f6p	D-Fructose 6-phosphate	−0.0709
g3p	Glyceraldehyde 3-phosphate	−0.129
g6p	D-Glucose 6-phosphate	−0.205
gln-L	L-Glutamine	−0.2557
glu-L	L-Glutamate	−4.9414
h	H^+	59.81
h2o	H_2O	−59.81
nad	Nicotinamide adenine dinucleotide (NAD^+)	−3.547
nadh	Nicotinamide adenine dinucleotide-reduced	3.547
nadp	Nicotinamide adenine dinucleotide phosphate ($NADP^+$)	13.0279
nadph	Nicotinamide adenine dinucleotide phosphate-reduced	−13.0279
oaa	Oxaloacetate	−1.7867
pep	Phosphoenolpyruvate	−0.5191
pi	Phosphate	59.81
pyr	Pyruvate	−2.8328
r5p	α-D-Ribose 5-phosphate	−0.8977

[a] In the biomass reaction, 23 different metabolites are consumed or produced to simulate growth. The metabolites that are consumed have negative stoichiometric coefficients, and the metabolites that are produced have positive coefficients.

are precursor metabolites. For example, the amino acid L-alanine is synthesized from pyruvate and L-glutamate, so both of these metabolites are consumed in the biomass reaction. Several metabolites are actually produced by the biomass reaction. ADP, protons, and inorganic phosphate are produced by the hydrolysis of ATP in the balanced reaction "atp + h2o → adp + h + pi." 2-Oxoglutarate is produced during the synthesis of amino acids, when L-glutamate transfers its amino group to another compound in a transamination reaction. Coenzyme A is produced when acetyl-CoA is consumed, and NAD^+ is reduced to NADH and NADPH is oxidized to $NADP^+$ during biomass synthesis.

Additional energetic requirements exist for growth beyond what is needed to generate the macromolecular content of the cell. These energetic maintenance requirements are for growth-associated maintenance (e.g., protein polymerization costs) and non-growth-associated maintenance (e.g., membrane leakage). To represent growth-associated maintenance, ATP is converted to ADP at 59.81 mmol gDW^{-1} h^{-1}, accounting for energy used in cell division and other growth processes. Non-growth-associated maintenance is not part of the biomass reaction. Instead, it is represented with a lower bound of 8.39 mmol gDW^{-1} h^{-1} on the ATP maintenance reaction (*ATPM*), simulating energy used for protein turnover and other processes that do not change with growth.

BOOLEAN CORE *E. COLI* TRANSCRIPTIONAL REGULATION

In addition to the metabolic reconstruction, the core *E. coli* model also contains a Boolean representation of part of the associated transcriptional regulatory network. This network is a modified subset of the genome-scale transcriptional regulatory reconstruction *i*MC1010 (58). In response to external and internal stimuli, in silico transcription factors either activate or repress genes associated with metabolic reactions. This regulation improves the predictive fidelity of the metabolic model by imposing additional context-dependent constraints on certain reactions. The transcriptional regulatory reconstruction consists of a set of Boolean rules that dictate whether a gene is either fully induced or fully repressed. If the genes associated with an enzyme or transport protein/complex are repressed, then in silico flux is constrained to zero for the corresponding reaction. The solution space of the network shrinks when these additional constraints are imposed. Reactions that are not used because of regulatory effects are thus restricted, so when using flux balance analysis, the optimal flux distribution will be consistent with known regulation. This optimal flux distribution may be different from the flux distribution of an unregulated model. In this case, the flux distribution of the unregulated model violated at least one regulatory constraint, making it biologically unrealistic. The use of computationally implemented Boolean rules in a genome-scale model has been shown to lead to more accurate flux balance analysis predictions (58).

A gene is considered to be induced when evaluation of the corresponding Boolean rule gives "true." In contrast,

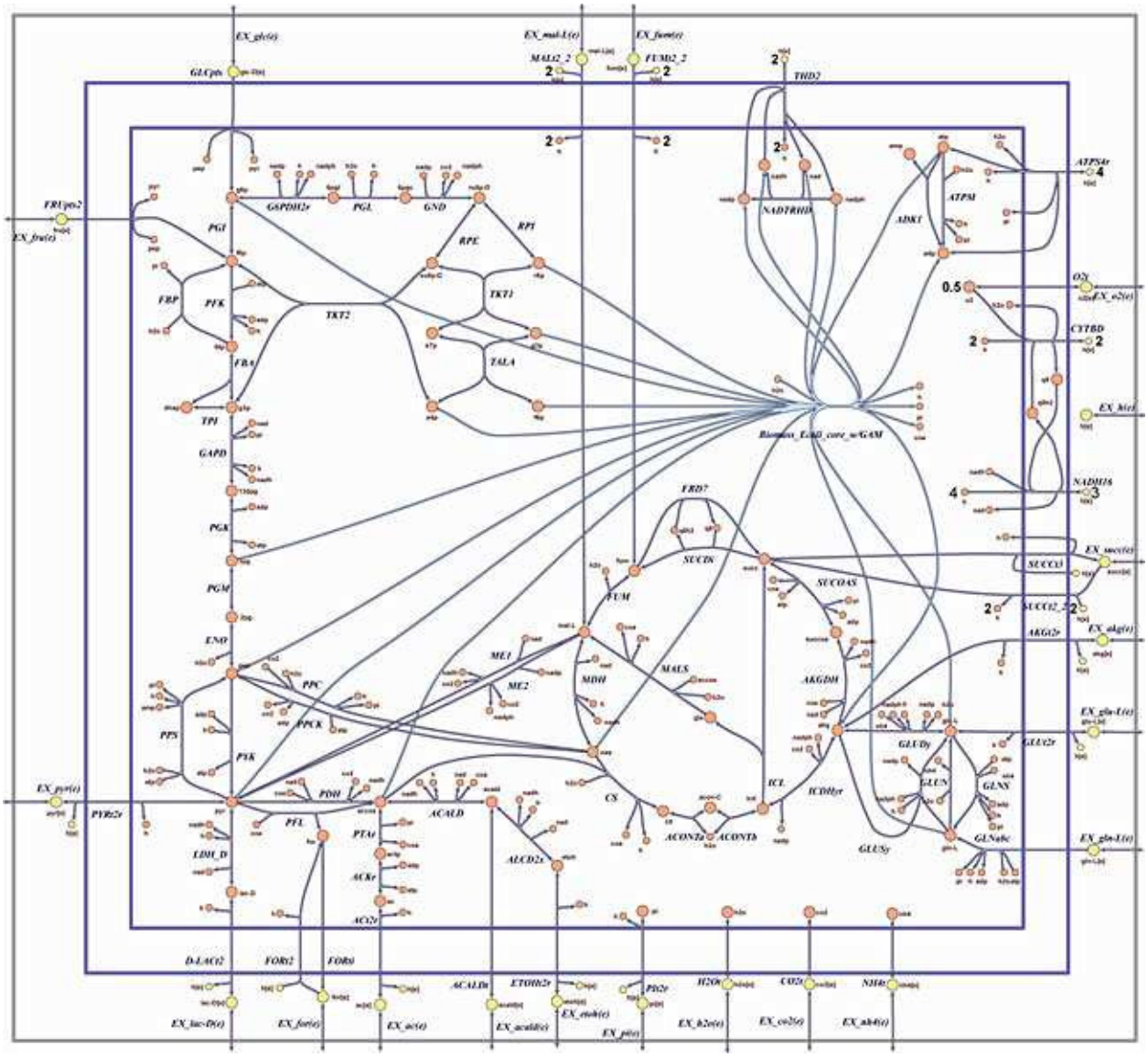

Figure 13 The biomass reaction (light blue) of the *E. coli* core model shown in the context of the entire model. This is a single reaction that consumes and produces metabolites from many different pathways in the model. In order to grow, *E. coli* must be able to produce all of the compounds that are consumed. The metabolites that are consumed enter the reaction from the left, and the by-products that are produced leave the reaction to the right. The biomass reaction drains the precursors and energy/redox carriers 3-phospho-D-glycerate, acetyl-CoA, ATP, D-erythrose 4-phosphate, D-fructose 6-phosphate, glyceraldehyde 3-phosphate, D-glucose 6-phosphate, L-glutamine, L-glutamate, H_2O, NAD^+, NADPH, oxaloacetate, phosphoenolpyruvate, pyruvate, and α-D-ribose 5-phosphate from the network while producing ADP, 2-oxoglutarate, coenzyme A, H^+, NADH, $NADP^+$, and phosphate.

a gene is considered to be repressed when evaluation of the corresponding Boolean rule gives "false." Boolean logic is used to evaluate each Boolean rule. For example, consider the enzyme phosphoenolpyruvate synthase, *PPS*, which catalyzes the first step of gluconeogenesis, the conversion of pyruvate to phosphoenolpyruvate. The gene for phosphoenolpyruvate synthase is *pps* and its Boolean rule is simply "FruR." That is, if FruR is "true," then the *pps* gene is induced allowing in silico flux through the reaction catalyzed by phosphoenolpyruvate synthase, *PPS*. FruR is a transcriptional regulator that is active when the cytoplasmic concentration of D-fructose 1,6-bisphosphate, *FDP*, is low (160). The Boolean rule for FruR is "NOT *surplusFDP*." That is, if there is no surplus of D-fructose 1,6-bisphosphate, then FruR is "true," and therefore the *pps* gene is induced, allowing in silico gluconeogenic flux through the reaction catalyzed by phosphoenolpyruvate synthase, *PPS*. In contrast, if

Table 16 Regulatory rules for metabolic genes in the core model

Gene	Locus	Reaction abbr.	Regulatory rule
aceA	b4015	*ICL*	(NOT IclR) AND ((NOT ArcA) OR FruR)
aceB	b4014	*MALS*	(NOT IclR) AND ((NOT ArcA) OR FruR)
aceE	b0114	*PDH*	(NOT PdhR) OR Fis
aceF	b0115	*PDH*	(NOT PdhR) OR Fis
adhE	b1241	*ACALD* and *ALCD2x*	(NOT o2[e]) OR (NOT (o2[e] AND FruR)) OR Fis
cydA	b0733	*CYTBD*	(NOT Fnr) OR ArcA
cydB	b0734	*CYTBD*	(NOT Fnr) OR ArcA
dctA	b3528	*FUMt2_2, MALt2_2*, and *SUCCt2_2*	*CRPnoGLM* AND (NOT ArcA) AND DcuR
focA	b0904	*FORt2* and *FORti*	ArcA OR (Fnr AND *CRPnoGLC*)
focB	b2492	*FORt2* and *FORti*	ArcA OR (Fnr AND *CRPnoGLC*)
frdA	b4154	*FRD7*	Fnr OR DcuR
frdB	b4153	*FRD7*	Fnr OR DcuR
frdC	b4152	*FRD7*	Fnr OR DcuR
frdD	b4151	*FRD7*	Fnr OR DcuR
fumA	b1612	*FUM*	NOT (ArcA OR Fnr)
fumB	b4122	*FUM*	Fnr OR *CRPnoGLC* OR DcuR
fumC	b1611	*FUM*	NOT ArcA
gdhA	b1761	*GLUDy*	NOT (Nac OR glu-L[e])
glcA	b2975	*D_LACt2*	(NOT ArcA) AND GlcC
glcB	b2976	*MALS*	(NOT ArcA) AND GlcC
glnA	b3870	*GLNS*	*CRPnoGLC*
gltB	b3212	*GLUSy*	(NOT (*NRI_hi* AND glu-L[e]))
gltD	b3213	*GLUSy*	(NOT (*NRI_hi* AND glu-L[e]))
lldP	b3603	*D_LACt2*	NOT ArcA
manX	b1817	*FRUpts2* and *GLCpts*	*CRPnoGLM* OR (NOT Mlc)
manY	b1818	*FRUpts2* and *GLCpts*	*CRPnoGLM* OR (NOT Mlc)
manZ	b1819	*FRUpts2* and *GLCpts*	*CRPnoGLM* OR (NOT Mlc)
mdh	b3236	*MDH*	NOT ArcA
nuoA	b2288	*NADH16*	(NOT (ArcA OR Fnr))
nuoB	b2287	*NADH16*	(NOT (ArcA OR Fnr))
nuoC	b2286	*NADH16*	(NOT (ArcA OR Fnr))
nuoE-N	b2276-b2285	*NADH16*	(NOT (ArcA OR Fnr))
pflA	b0902	*PFL*	ArcA OR (Fnr AND *CRPnoGLC*)
pflB	b0903	*PFL*	ArcA OR (Fnr AND *CRPnoGLC*)
pflC	b3952	*PFL*	ArcA OR Fnr
pflD	b3951	*PFL*	ArcA OR Fnr
pitB	b2987	*PIt2r*	NOT PhoB
pps	b1702	*PPS*	FruR
ptsG	b1101	*GLCpts*	(NOT Mlc) OR (NOT FruR)
pykF	b1676	*PYK*	NOT FruR
sdhA	b0723	*SUCDi*	(NOT (ArcA OR Fnr)) OR *CRPnoGLC* OR Fis

(continued)

Table 16 Regulatory rules for metabolic genes in the core model *(continued)*

Gene	Locus	Reaction abbr.	Regulatory rule
sdhB	b0724	*SUCDi*	(NOT (ArcA OR Fnr)) OR *CRPnoGLC* OR Fis
sdhC	b0721	*SUCDi*	(NOT (ArcA OR Fnr)) OR *CRPnoGLC* OR Fis
sdhD	b0722	*SUCDi*	(NOT (ArcA OR Fnr)) OR *CRPnoGLC* OR Fis
tdcD	b3115	*ACKr*	*CRPnoGLC* OR Fnr
tdcE	b3114	*PFL*	*CRPnoGLC* OR Fnr
yneH	b1524	*GLUN*	(NOT glc-D[e]) OR (nh4[e] AND (NOT *CRPnoGLC*))

surplusFDP is "true," then FruR is "false," and therefore *pps* is repressed.

Regulatory conditions, such as *surplusFDP*, are variables that represent a complex regulatory rule for a transcription factor that cannot be accurately represented with only one variable. The regulatory rule for *surplusFDP* is "((NOT *FBP*) AND (NOT (*TKT2* OR *TALA* OR *PGI*))) OR fru[e]." If fru[e] is "true," then *surplusFDP* is "true," independent of the state of the other variables. If fructose-bisphosphatase, *FBP*, is "false," and any one of transketolase, *TKT2*, transaldolase, *TALA*, or glucose-6-phosphate isomerase, *PGI*, is "false," then *surplusFDP* is "true"; therefore, FruR is "false" and *pps* is repressed. By using Boolean logic, all rules in a regulatory network can be reduced to either "true" or "false," and ultimately this dictates whether each metabolic gene is induced or repressed. Not every gene in the metabolic network is controlled by the regulatory network, so the unregulated genes are assumed to always be active, and their fluxes are never constrained to zero. Table 16 lists the Boolean regulatory rules for regulated metabolic genes, and Table 17 lists the Boolean regulatory rules for transcription factors and regulatory conditions in the transcriptional regulatory network. An abstract overview of part of the regulatory network is depicted in Fig. 14.

Regulation of Anoxic Growth

Under anoxic conditions, the transcription factors ArcA and Fnr act as global regulators that induce many different genes needed for fermentation and growth without oxygen. However, the principal function of ArcA and Fnr is to repress genes not required when oxygen is abundant (161). When oxygen availability is reduced, the drop in redox potential signals the phosphorylation and thereby activation of the global regulator ArcA. Only in anaerobic conditions is the global transcriptional regulator Fnr also activated. When there is oxygen in the cytoplasm, an oxidized 4Fe-4S cluster in Fnr inactivates it. The regulatory rule "NOT o2[e]" is the same for ArcA and Fnr. Therefore, when o2[e] is false, ArcA and Fnr are true, representing activation.

Both ArcA and Fnr induce fermentative genes such as *pflA*, *pflB*, *pflD*, and *pflC*, coding for pyruvate formate lyase, *PFL* (162), and *focA* and *focB* coding for formate

Table 17 Regulatory rules for transcriptional regulators and regulatory conditions

Regulator	Locus	Regulatory rule
ArcA	b4401	NOT o2[e]
DcuR	b4124	DcuS
DcuS	b4125	succ[e] OR fum[e] OR mal-L[e]
FadR	b1187	glc-D[e] OR (NOT ac[e])
Fis	b3261	*Biomass_Ecoli_core_w_GAM*
Fnr	b1334	NOT o2[e]
FruR	b0080	NOT *surplusFDP*
GlcC	b2980	ac[e]
GlnG	b3868	NOT nh4[e]
IclR	b4018	FadR
Mlc	b1594	NOT glc-D[e]
Nac	b1988	*NRI_low*
PdhR	b0113	NOT *surplusPYR*
PhoB	b0399	PhoR
PhoR	b0400	NOT pi[e]
CRPnoGLC	b3357	NOT glc-D[e]
CRPnoGLM	b3357	NOT (glc-D[e] OR mal-L[e] OR lac-D[e])
NRI_hi		*NRI_low*
NRI_low		GlnG
surplusFDP		((NOT *FBP*) AND (NOT (*TKT2* OR *TALA* OR *PGI*))) OR fru[e]
surplusPYR		(NOT (*ME2* OR *ME1*)) AND (NOT (*GLCpts* OR *PYK* OR *PFK* OR *LDH_D* OR *SUCCt2_2*))

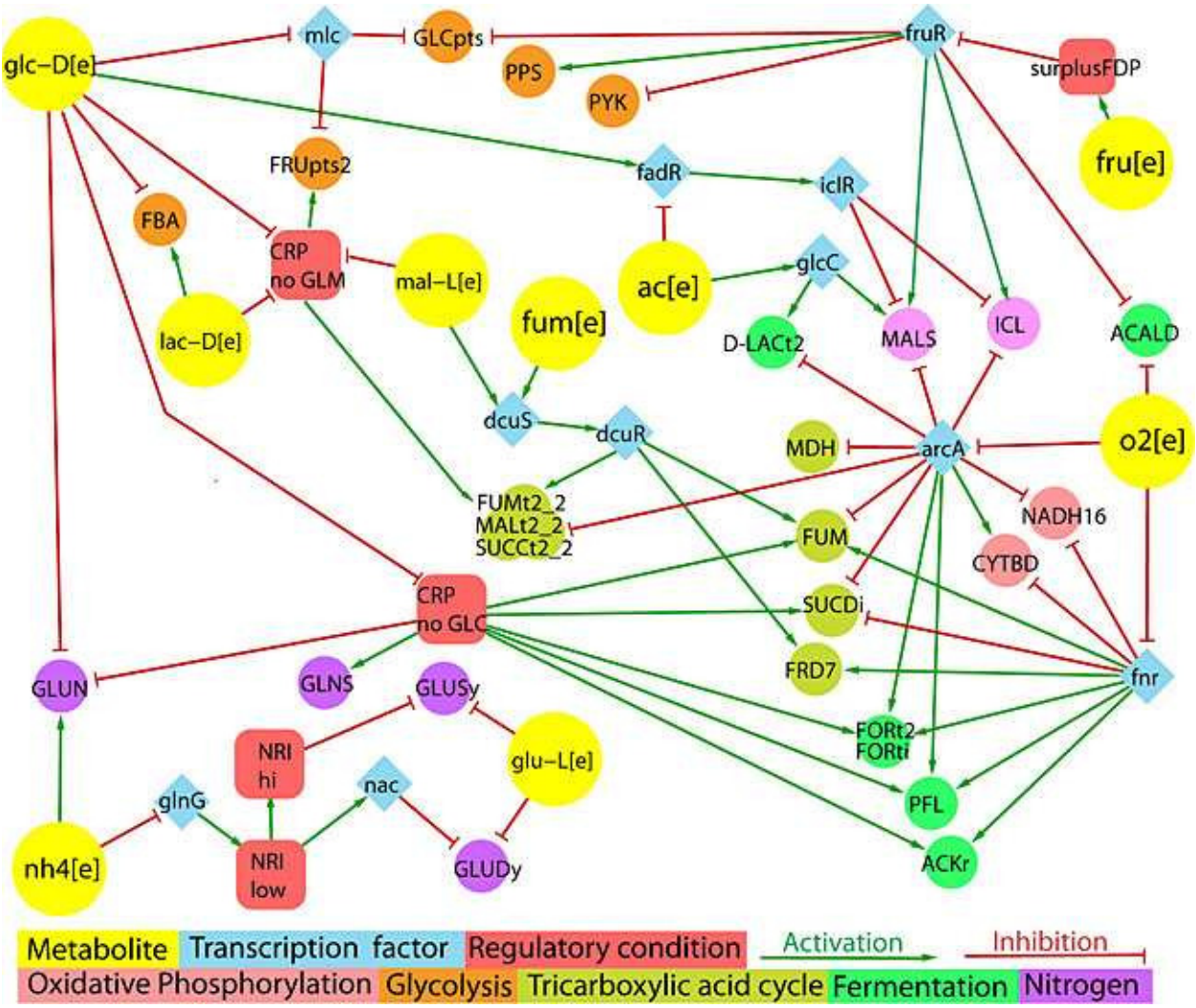

Figure 14 A schematic overview of part of the *E. coli* core regulatory network from an environmental stimulus-response perspective. For the sake of clarity, reaction abbreviations are displayed rather than the gene(s) that code for the associated protein or protein complex. The Boolean state of in silico reactions, true = flux, or false = zero flux, may be determined by traversing the activation or inhibition links, beginning at an environmental metabolite stimulus (yellow circles). An anoxic environment is represented by o2[e] = false. Oxygen inactivates Fnr, represented by an inhibition link from o2[e] to Fnr. When o2[e] is false, then Fnr is true, indicating that Fnr is active in the absence of oxygen. Active Fnr then induces or represses the expression of genes for various reactions. Induction of the genes for fumarase, *FUM*, by Fnr, is represented by an activation link from Fnr to *FUM*. Repression of the genes for oxidative phosphorylation by Fnr is represented by the inactivation links from Fnr to NADH dehydrogenase, *NADH16*, and cytochrome oxidase, *CYTBD*, i.e., activation assigns the target node the same Boolean status as the source node, whereas inhibition sets the Boolean status of the target to the opposite of the source. The central role of the global transcriptional regulators ArcA, Fnr, Crp, and FruR is evident from the number of reactions under their control. Boolean regulatory rules also encode combinatorial regulation of the same gene by multiple transcription factors. For details, see Table 16.

transport, *FORt2* and *FORti*. The regulatory rule for *pflC* is "ArcA or Fnr." Therefore, if either ArcA or Fnr are true, then *pflC* is true, representing expression. Fnr also induces *tdcD*, coding for the fermentative enzyme acetate kinase, *ACKr*. Anoxic conditions also induce *adhE*, encoding the enzymes for both acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x*. The regulatory rule for *adhE* is "(NOT o2[e]) OR (NOT (o2[e] AND FruR)) OR Fis," representing its indirect induction when oxygen is absent or derepression by the inactivity of the transcriptional regulator FruR. Fis is discussed in "Regulation by Fis."

In the absence of a common electron acceptor such as oxygen or nitrate, the reactions of the TCA cycle no longer operate as an energy-producing cycle. Instead,

they function as two separate biosynthetic pathways. Beginning at oxaloacetate, a reductive pathway via fumarate reductase, *FRD7*, producing succinyl-CoA, and an oxidative pathway producing 2-oxoglutarate. In the absence of oxygen, ArcA represses a number of genes in the TCA cycle that are unnecessary for its cyclic operation, including *sdhA*, *sdhB*, *sdhC*, and *sdhD*, coding for succinate dehydrogenase (irreversible), *SUCDi* (75, 163), in addition to *fumA* and *fumC* coding for fumarase, *FUM* (164), and *mdh* coding for malate dehydrogenase, *MDH*. The regulatory rule for *sdhA-D*, "(NOT (ArcA OR Fnr)) OR Crp OR Fis," indicates that these genes are expressed when ArcA and Fnr are false, or when either of the transcriptional regulators Crp or Fis are true.

The cell adapts to changing environmental oxygen conditions by utilizing different isozymes of fumarase, *FUM*. Both Fnr and ArcA repress *fumA* coding for fumarase A (164), but Fnr induces *fumB* coding for the isozyme fumarase B that has greater affinity for malate as a substrate (165). The regulatory rule for *fumA* is "NOT (ArcA OR Fnr)," so the *fumA* isozyme is expressed when ArcA and Fnr are false. The regulatory rule for *fumB* is "Fnr OR Crp OR DcuR," so the *fumB* isozyme is expressed when any one of Fnr, Crp, or DcuR is true. The transcriptional regulator DcuR is discussed in "Growth on Acetate or C_4-Dicarboxylate Compounds," below. Fnr complements induction by DcuR of *frdA*, *frdB*, *frdC*, and *frdD*, coding for fumarate reductase, *FRD7* (166). In anoxic conditions, fumarate reductase, *FRD7*, catalyzes the reverse reaction to that catalyzed by succinate dehydrogenase, *SUCDi*, in oxic conditions. The regulatory rule for *frdA-D* is "Fnr OR DcuR," meaning that these genes are expressed when either Fnr or DcuR are activated.

Both ArcA and Fnr downregulate oxidative phosphorylation by repressing the *nuoA-nuoN* operon coding for NADH dehydrogenase (ubiquinone-8: & 3 protons), *NADH16* (167, 168). This reaction is used for aerobic respiration and is therefore unnecessary in anoxic conditions. The corresponding regulatory rule for *nuoA-N* is "(NOT (ArcA OR Fnr))." When oxygen is present at low concentration ArcA induces *cydA* and *cydB*, coding for the reaction cytochrome oxidase bd (ubiquinol-8: -2 protons), *CYTBD* (166). In fully anoxic conditions these same genes *cydA* and *cydB* coding for cytochrome oxidase bd (ubiquinol-8: 2 protons), *CYTBD*, are repressed by Fnr (166). The regulatory rule for *cydA* and *cydB* is "(NOT Fnr) OR ArcA." ArcA downregulates the glyoxylate cycle by repressing *aceB* and *glcB*, coding for malate synthase, *MALS*, and *aceA* coding for isocitrate lyase, *ICL* (169). ArcA also represses a number of transporters, including *glcA* and *lldP* coding for D-lactate transport via proton symport, *D_LACt2*, and *dctA* coding for the aerobic transporter for fumarate, malate, and succinate, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2* (170). Many of the latter genes are also regulated by other transcription factors giving rise to a combinatorial expansion in the total number of network states.

Catabolite Repression

In a medium containing glucose and another substrate such as lactate or malate, *E. coli* preferentially catabolizes glucose until it is depleted, thereafter switching to catabolism of the less desirable substrate (171). The repression of enzymes for catabolism of a less desirable substrate by the presence of a desirable substrate is generally termed catabolite repression (172) (see Fig. 15). Cytoplasmic concentrations of the cofactor cyclic adenosine monophosphate, cAMP, allosterically control the activity of the cAMP receptor protein, Crp (173). The presence of glucose reduces the cytoplasmic concentration of Crp-cAMP that is necessary for induction of various genes necessary for catabolism of less desirable substrates. To represent regulation by Crp-cAMP in a Boolean manner, extra regulatory condition variables are necessary to represent Crp-cAMP under different conditions. The *CRPnoGLM* regulatory condition is false when either glucose (glc-D[e]), malate (mal-L[e]), or lactate (lac-D[e]) are present in the media. In this condition the gene *dctA* is repressed, which codes for the fumarate, malate, and succinate transporters, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2* (170). The regulatory rule for *dctA* is "*CRPnoGLM* AND (NOT ArcA) AND DcuR." Therefore ArcA must also be false and DcuR must also be true for *dctA* to be expressed. When the *CRPnoGLM* regulatory condition is false, the genes *manX*, *manY*, and *manZ* are repressed downregulating fructose transport via PEP:Pyr PTS, *FRUpts2*. Note that *manX*, *manY*, and *manZ* can also code for subunits of D-glucose transport via PEP:Pyr PTS, *GLCpts*. The latter can also be encoded by *ptsH*, *ptsI*, *malX*, *crr*, and *ptsG*, providing an independent means of glucose transport.

When glucose is absent, the regulatory condition *CRPnoGLC* is true, representing activation of the global regulator Crp-cAMP. *CRPnoGLC* induces the genes *sdhA*, *sdhB*, *sdhC*, and *sdhD*, coding for succinate dehydrogenase, *SUCDi* (75). *CRPnoGLC* also upregulates

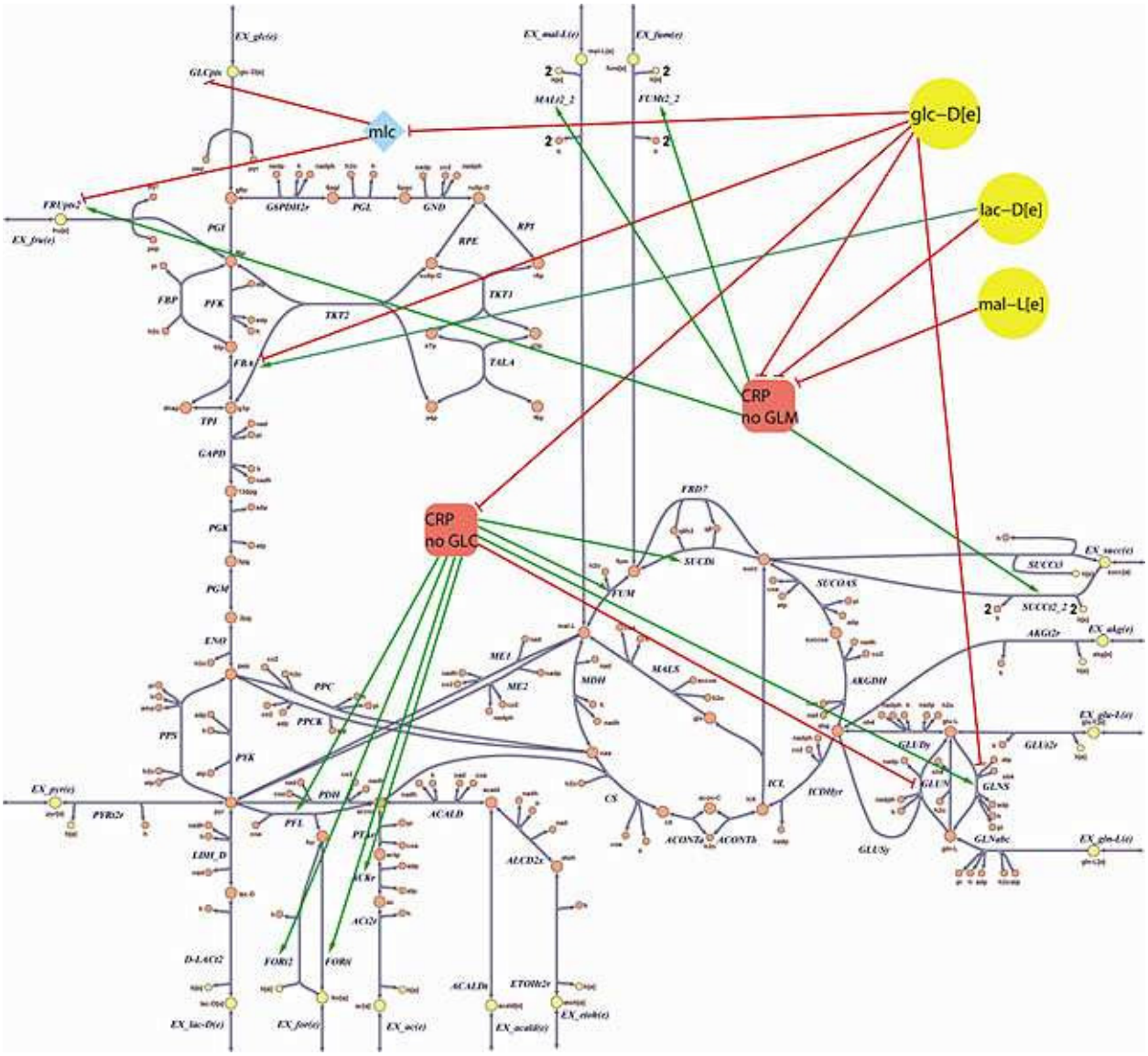

Figure 15 Overview of the effects of catabolite repression on the core *E. coli* model. When the metabolites glucose, (glc-D[e]), lactate (lac-D[e]), or malate (mal-L[e]) are absent from the media, the conditions *CRPnoGLM* and *CRPnoGLC* are activated, regulating a variety of metabolic reactions.

fermentation by inducing *pflA*, *pflB*, and *tdcE* coding for pyruvate formate lyase, *PFL*, inducing *focA* and *focB* coding for formate transport, *FORt2* and *FORti*, inducing *tdcD* coding for acetate kinase, *ACKr* (174). The regulatory rule for *focA* and *focB* is "ArcA OR (Fnr AND *CRPnoGLC*)"; therefore, these genes are induced when ArcA is true or when both Fnr and *CRPnoGLC* are true. When *CRPnoGLC* is true, this induces *fumB* coding for fumarase, *FUM*, in the TCA cycle. *CRPnoGLC* represses *glnA* coding for glutaminase, *GLUN*, but induces *yneH* coding for glutamine synthase, *GLNS*. Another transcription factor that is activated when glucose is absent is Mlc. Mlc represses *ptsG*, *manX*, *manY*, and *manZ*, which all code for subunits of the glucose and fructose phosphotransferase systems, *FRUpts2* and *GLCpts* (175). The regulatory rule for *manX-Z* is "*CRPnoGLM* OR (NOT Mlc)," indicating that these transport genes are induced when *CRPnoGLM* is true or Mlc is false.

Growth on Acetate or C_4-Dicarboxylate Compounds

The gene *glcC* codes for a transcription factor that is induced when acetate (ac[e]) is present in the media (169). The regulatory rule for GlcC is therefore simply

"ac[e]." GlcC induces *glcA* and *glcB*, which code for D-lactate transport via proton symport, *D_LACt2*, and malate synthase, *MALS*, respectively (169). The regulatory rule for *glcA* and *glcB* is "(NOT ArcA) AND GlcC," so these genes are induced when acetate and oxygen are present. The dual transcriptional regulator FadR is activated by the presence of glucose or the absence of acetate in the media, and it is a regulator of fatty acid synthesis. FadR activates transcription of the gene *iclR*, which is also a transcription factor (176). IclR represses transcription of the genes *aceB* and *aceA* (177), which code for the malate synthase, *MALS*, and isocitrate lyase, *ICL*, enzymes, respectively. The overall effect of *fadR* is thus to inhibit the glyoxylate cycle when the cell is not consuming acetate. The genes *aceB* and *aceA* have the regulatory rule "(NOT IclR) AND ((NOT ArcA) OR FruR)," indicating that the glyoxylate cycle is active when glucose and oxygen are present or when glucose is present and acetate is absent.

The response to C_4-dicarboxylate compounds is regulated by the two-component system *dcuR* and *dcuS*. *dcuS* codes for a sensor histidine kinase that is activated when fumarate (fum[e]), L-malate (mal-L[e]), or succinate (succ[e]) are present in the media (178). Once activated, the DcuS protein phosphorylates DcuR, activating it. DcuR then induces the gene *dctA* (170), which codes for the transporter for fumarate, malate, and succinate, *FUMt2_2*, *MALt2_2*, and *SUCCt2_2*, allowing these metabolites to be consumed. DcuS also upregulates *fumB* (165), which codes for the fumarase B isozyme. Finally, DcuS upregulates *frdA*, *frdB*, *frdC*, and *frdD* (178, 166), which are subunits of the fumarate reductase enzyme, *FRD7*. These two reactions are responsible for interconversion of succinate, fumarate, and malate. The regulatory rule for *frdA-D* is "Fnr OR DcuR," indicating that either when oxygen is present or fumarate, L-malate, or succinate is present in the media, then fumarate reductase is induced.

Glycolysis versus Gluconeogenesis

When cytoplasmic concentrations of D-fructose 1,6-bisphosphate (fdp) are low, the dual transcriptional regulator FruR represses glycolytic and fermentative enzymes, yet simultaneously induces gluconeogenic genes (160). In contrast, surplus D-fructose 1,6-bisphosphate binds to FruR, dislodging it from its binding sites and thereby derepressing glycolytic and fermentative enzymes and deactivating gluconeogenic enzymes. Since current constraint-based models do not explicitly model metabolite concentrations, the regulatory condition *surplusFDP* is used to indicate conditions of excess D-fructose 1,6-bisphosphate. The *surplusFDP* condition is met when fructose (fru[e]) is present in the media or the reactions *FBP* and any of *TKT2*, *TALA*, or *PGI* have zero flux. When *surplusFDP* is false, then FruR is true, thereby repressing glycolytic and fermentative enzymes and inducing gluconeogenic enzymes.

FruR induces *ptsG* coding for D-glucose transport via PEP:Pyr PTS, *GLCpts*, but represses *pykF* coding for the glycolytic enzyme pyruvate kinase, *PYK* (179). FruR also downregulates fermentation by repressing *adhE*, coding for acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x* (180, 181). FruR upregulates the glyoxylate cycle by inducing *aceA* coding for isocitrate lyase, *ICL*, and inducing *aceB* coding for malate synthase, *MALS* (177). FruR also upregulates gluconeogenesis by inducing *pps*, which encodes phosphoenolpyruvate synthase *PPS* (182). In summary, FruR is capable of reversing the flow of carbon to replenish glycolytic intermediates as sensed by the level of D-fructose 1,6-bisphosphate.

The dual transcriptional regulator *pdhR* (pyruvate dehydrogenase complex regulator) downregulates pyruvate dehydrogenase when the pyruvate concentration in the cell is low (183). The Boolean regulatory rule for PdhR is "NOT *surplusPYR*" (see Fig. 17). High pyruvate concentration is represented by the variable *surplusPYR*, which is true when there is no flux through *ME1* or *ME2*, and no flux through either one of *GLCpts*, *PYK*, *PFK*, *LDH_D*, or *SUCCt2_2*. The Boolean rule for *surplusPYR* is "(NOT (*ME2* OR *ME1*)) AND (NOT (*GLCpts* OR *PYK* OR *PFK* OR *LDH_D* OR *SUCCt2_2*))." PdhR inhibits the genes *aceE* and *aceF* that both code for subunits of the pyruvate dehydrogenase complex, *PDH*. The regulatory rule for *aceE* and *aceF* is "(NOT PdhR) OR Fis," meaning that these genes are repressed when PdhR is true and meaning that pyruvate dehydrogenase is repressed when cytoplasmic pyruvate concentration is low.

Regulation by Fis

The DNA-binding protein Fis regulates the expression of many genes by bending the genomic DNA, changing its topological structure. It has been shown to directly or indirectly regulate 21% of the genes in *E. coli*, and 894 Fis-associated regions of the genome have been identified

(63). Transcription of *fis* is activated when the cell is in exponential growth phase. Fis then activates the genes *aceE*, *aceF*, coding for two of the subunits of the pyruvate dehydrogenase enzyme, *PDH*. It also activates transcription of the succinate dehydrogenase, *SUCDi*, subunits *sdhC*, *sdhD*, *sdhA*, and *sdhB* (75). Fis can also activate *adhE*, which codes for an enzyme that carries out both the acetaldehyde dehydrogenase (acetylating), *ACALD*, and alcohol dehydrogenase (ethanol), *ALCD2x*, reactions (180). When modeling the balanced steady-state growth typical of exponential phase, the state of Fis is always set to true. Other growth phases can be modeled with the full *E. coli* Boolean regulatory model (58).

Nitrogen Regulation

The response to low nitrogen concentration in *E. coli* is a complex process (Fig. 16). There is a fast (low-level) response and a slower (high-level) response, represented in the Boolean regulatory model with the regulatory condition variables *NRI_low* and *NRI_hi*. The low-level response is activated by the transcription factor GlnG. GlnG is activated by low extracellular ammonium (nh4[e]) concentration and regulates nitrogen levels by induction or repression of many different genes (184). The high-level nitrogen response is activated by the low-level response, so *NRI_hi* is always activated after *NRI_low* is activated.

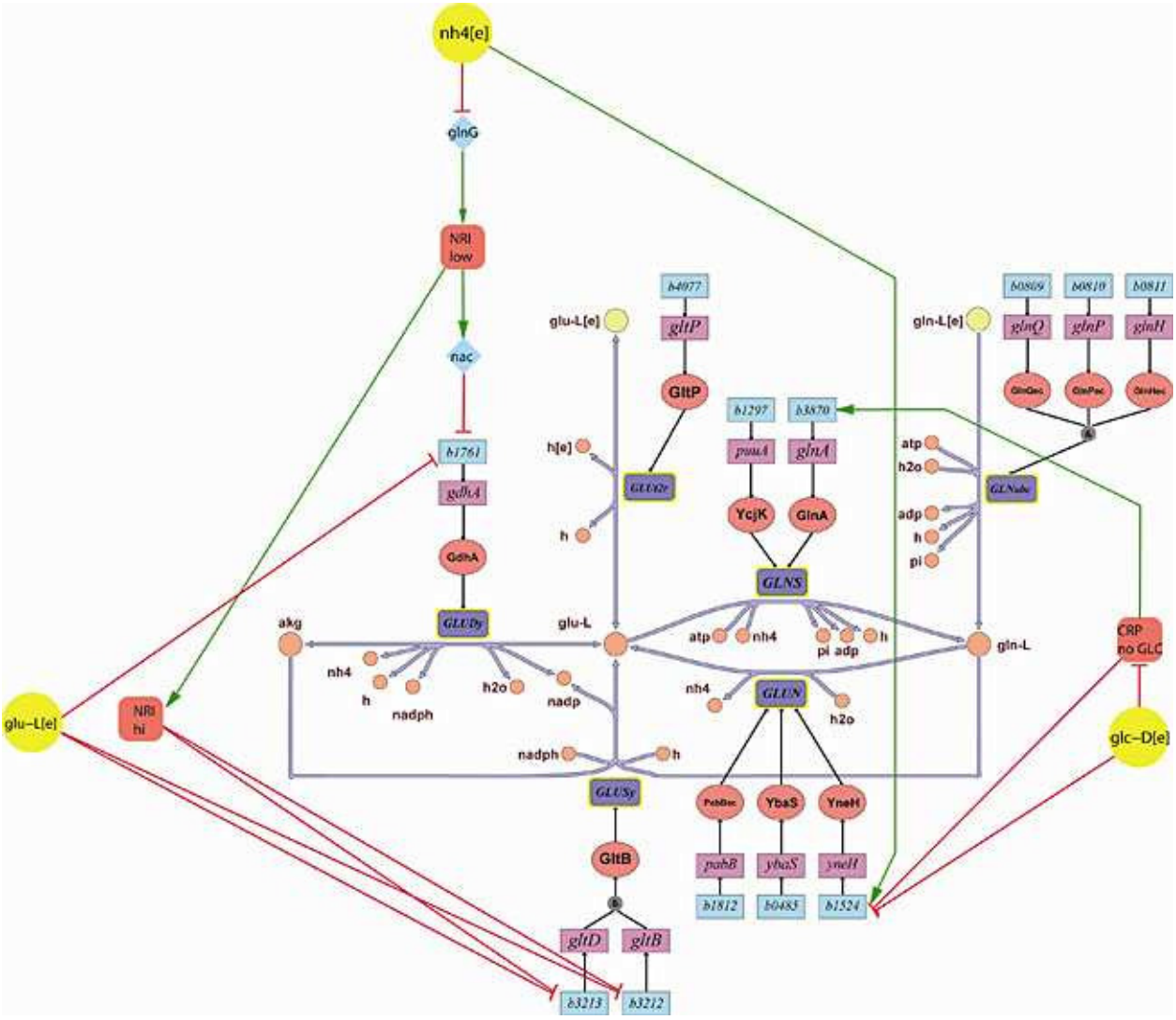

Figure 16 Regulation of nitrogen metabolism in the *E. coli* core model. Extracellular ammonium (nh4[e]) activates the low- and high-level nitrogen responses, *NRI_low* and *NRI_hi*, which, along with extracellular glutamate (glu-L[e]), inhibit the reactions glutamate dehydrogenase, *GLUDy*, and glutamate synthase, *GLUSy*. Glutaminase, *GLUN*, is also activated by extracellular ammonium. Extracellular glucose (glc-D[e]) through *CRPnoGLC* inhibits glutamine synthetase, *GLNS*, and glutaminase, *GLUN*.

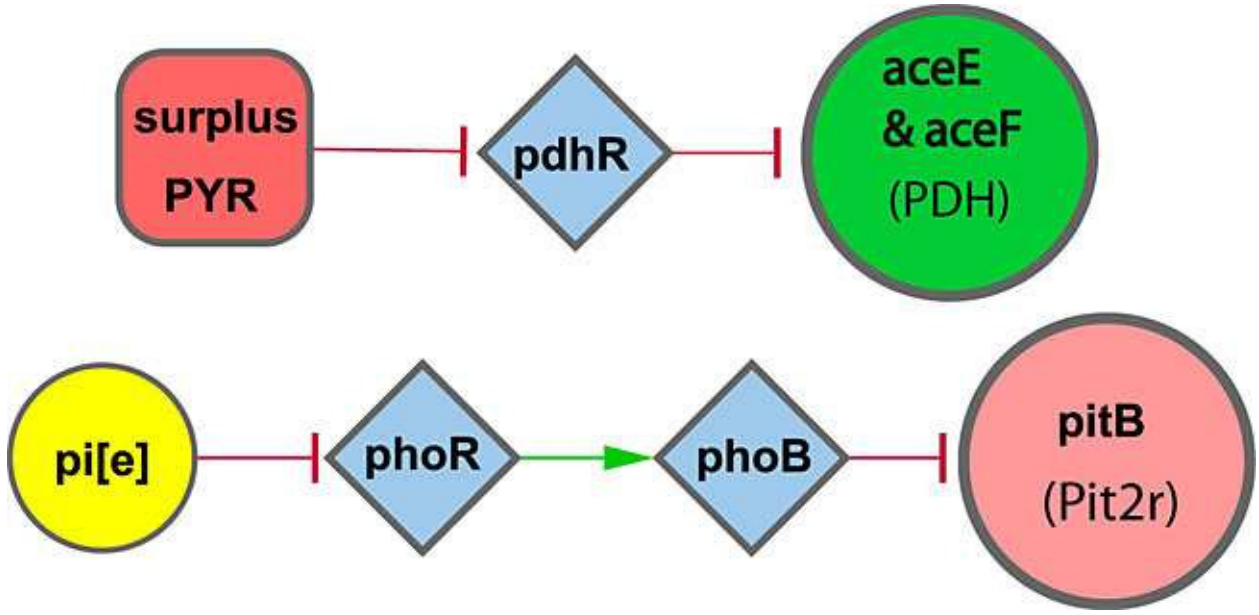

Figure 17 A schematic representation of cytoplasmic pyruvate concentration and phosphorus transport in the *E. coli* core model. Transcription factors are indicated by blue diamonds. The genes *aceE* and *aceF* both code for subunits of the pyruvate dehydrogenase complex, *PDH*. The *pitB* gene codes for the phosphate transporter, *PIt2r*.

As a whole, the low- and high-level nitrogen responses conserve nitrogen by decreasing glutamate production. *NRI_low* induces transcription of the transcription factor Nac (nitrogen assimilation control). The product of this gene then represses *gdhA* (185), which codes for the glutamate dehydrogenase enzyme, *GLUDy*, which produces glutamate from 2-oxoglutarate. *NRI_hi* represses the genes *gltB* and *gltD* (47), which code for the subunits of the glutamate synthase enzyme, *GLUSy*. This enzyme produces glutamate through an alternate mechanism, so *NRI_hi* leads to further nitrogen conservation.

Phosphorus Regulation

Phosphorus uptake is regulated by the two-component system *phoR*/*phoB* (186). *phoR* codes for a sensor kinase that is phosphorylated when extracellular inorganic phosphate (pi[e]) is not present. The phosphorylated enzyme is activated, and it phosphorylates the transcriptional regulator PhoB. Phosphorylated PhoB then represses the *pitB* gene, which codes for the phosphate transporter, *PIt2r*. As indicated in Fig. 17, the regulatory rule for *pitB* is "NOT PhoB"; therefore, *pitB* is true when *phoB* is false, and inorganic phosphate is present. The overall effect of phosphorus regulation is to downregulate the phosphate transport reaction, *PIt2r*, when no extracellular inorganic phosphate is present.

USES OF METABOLIC MODELS

Once an accurate metabolic reconstruction is converted into a computational model, it may be used for a growing number of applications. Such models have been utilized to address a broad spectrum of basic and practical applications in five main categories: studies of evolutionary processes, analysis of network properties, interpretation of phenotypic screens, model-directed discovery, and metabolic engineering (Fig. 18). Metabolic reconstructions are a common denominator in the systems analysis of metabolic functions. As evident from "Description of the Core *E. coli* Metabolic Reconstruction" and "Boolean Core *E. coli* Transcriptional Regulation," a wealth of biological knowledge is encoded in a manually curated network reconstruction. "Understanding Metabolic Capabilities" describes how this knowledge may then be used to predict capabilities of a network that emerge from the interaction of multiple components. Likewise, the knowledge encoded in a model may be used to study the process of bacterial evolution (158). Applications include the interpretation of experimental adaptive evolution (10), horizontal gene transfer (9, 10), and evolution to minimal metabolic networks (11). Network reconstructions also provide a context for integration of high-throughput data from multiple complementary experiments, as described in "High-Throughput Data Analysis." Discovery of the

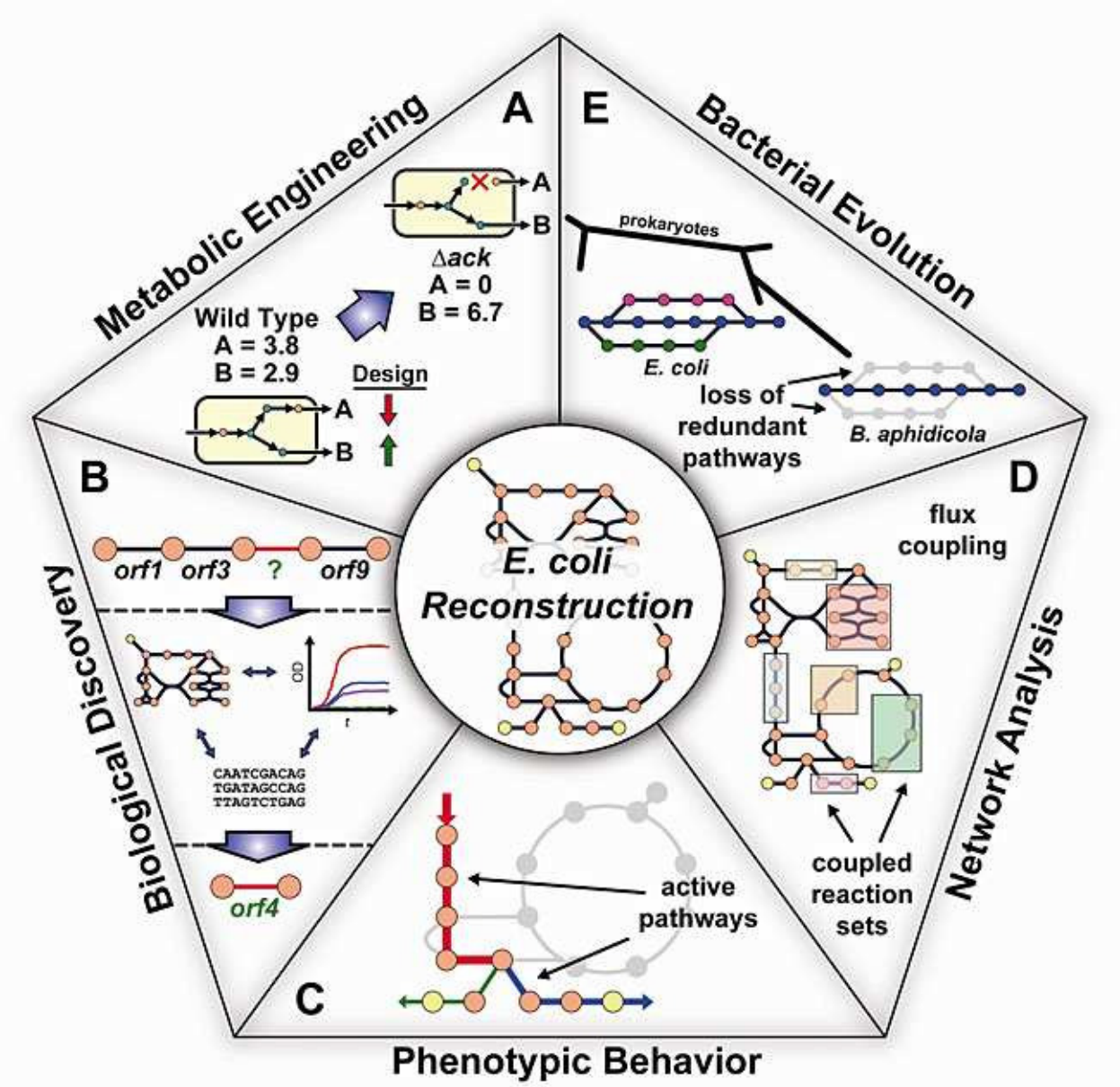

Figure 18 Metabolic models have practical applications in metabolic engineering, model-directed discovery, interpretations of phenotypic screens, analysis of network properties, and studies of evolutionary processes. (A) Metabolic engineering, e.g., predicting the effect from a loss-of-function mutation as part of in silico strain design to overproduce desired products. (B) Biological discovery involves prospective use of biochemical and genetic information included in the metabolic network along with additional data types to drive discovery, e.g., predicting genes responsible for orphan reactions. (C) Phenotypic studies, e.g., computational analyses of gene, metabolite, and reaction essentiality. (D) Network analysis, e.g., finding coupled reaction fluxes across different growth conditions. (E) Evolutionary studies have used metabolic models to interpret adaptive evolution events, horizontal gene transfer, and minimal metabolic network evolution. From reference 158.

biochemical function of previously uncharacterized genes using metabolic models is discussed in "Discovery." A growing application of genome-scale metabolic reconstructions is the prediction of optimal mutant strains in synthetic biological and engineering settings ("Synthetic Biology and Metabolic Engineering").

Understanding Metabolic Capabilities

Flux balance analysis and other constraint-based methods can be used to analyze the capabilities of a metabolic network (1). In flux balance analysis, the biomass reaction can be optimized (maximized) by use of linear programming software to simulate growth (50). The result is

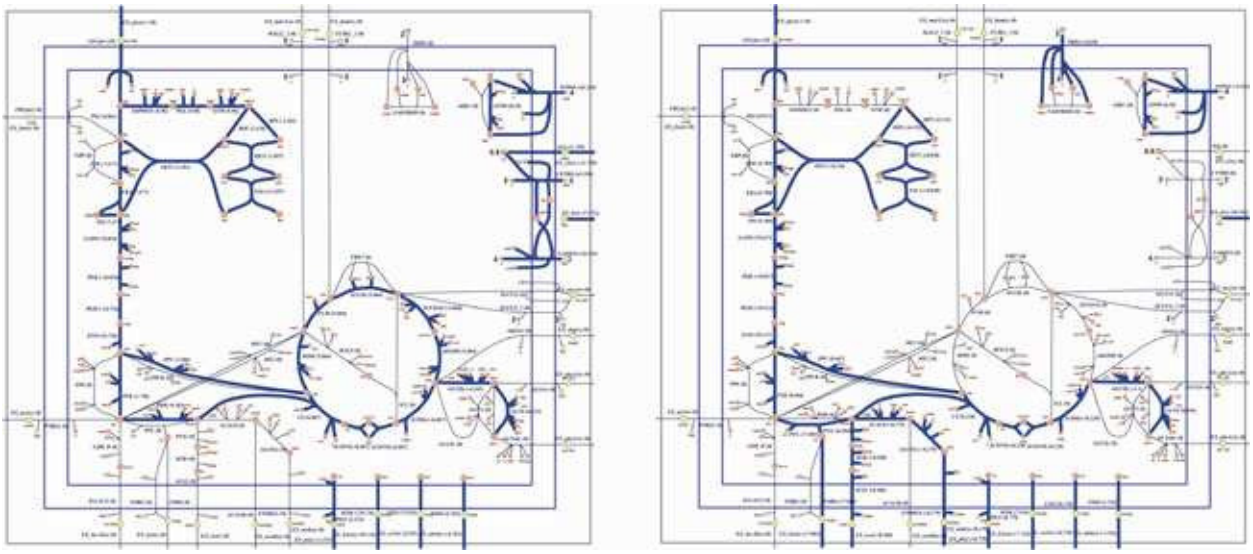

Figure 19 Flux distributions for growth on glucose under aerobic (left) and anaerobic (right) conditions. Reactions with thick blue arrows have nonzero fluxes, and reactions with thin black arrows carry zero flux.

an in silico prediction of steady-state flux through each reaction in the model including a prediction of the maximum balanced growth rate of the cell. Growth can be simulated under many different conditions, such as aerobic or anaerobic conditions, or growth on glucose or other substrates. Different conditions are simulated by changing the constraints on the exchange reactions. Exchange reactions, at the boundary of the model and the environment, act as sources or sinks of substrate and waste metabolites (Fig. 3). For example, anaerobic conditions are simulated when the lower bound of the O_2 exchange reaction, *EX_o2(e)*, is constrained to zero flux, allowing no O_2 to enter the system. To simulate different media, the exchange reactions for metabolites present in the media are constrained to have a lower bound equal to their desired uptake rate, and all metabolites not present will have exchange reactions with lower bounds constrained to zero.

Growth was simulated by using the core *E. coli* metabolic model (without regulation) with glucose as the only organic substrate, under both aerobic and anaerobic conditions; the resulting flux distributions are shown in Fig. 19. Under aerobic conditions with a glucose uptake rate (*EX_glc(e)*) of 10 mmol gDW^{-1} h^{-1}, the growth rate is 0.87 h^{-1}. The flux through the electron transport chain is high, and no organic by-products are secreted. Under anaerobic conditions with the same glucose uptake rate, the growth rate is 0.21 h^{-1}. The electron transport chain and most of the TCA cycle are not used at all, and formate, acetate, and ethanol are all secreted. When the regulatory model is combined with the metabolic model, the reactions *FORt2*, *FORti*, *FUMt2_2*, *ICL*, *MALS*, *MALt2_2*, *PFL*, and *SUCCt2_2* are inactivated under aerobic conditions when growing on glucose. However, since none of these reactions are used in the optimal flux distribution for growth on glucose, the growth rate is not affected. Under anaerobic conditions, the reactions *D_LACt2*, *FUMt2_2*, *ICL*, *MALS*, *MALt2_2*, *MDH*, *NADH16*, and *SUCCt2_2* are inactivated. These reactions are also not part of the optimal flux distribution, so anaerobic growth rate is still 0.21 h^{-1}. These results highlight the predictive capacity of assuming an optimal growth rate, given a particular environmental condition.

The maximum yields of important cofactors such as ATP, NADH, and NADPH can also be determined by

Table 18 Maximum cofactor production from glucose, aerobically

Cofactor	Yield	PPS,%[a]	Constraint
ATP	17.5	0	H^+
NADH	10	0	Energy, stoichiometry
NADPH	8.778	300	Energy, stoichiometry

[a] PPS is the percentage of the glucose uptake flux that enters the pentose phosphate shunt.

Table 19 Example COBRA Toolbox commands for performing flux balance analysis

Action	Command
Change bounds for anaerobic growth	model = changeRxnBounds (model,'EX_o2(e)',0,'l');
Change bounds for aerobic growth	model = changeRxnBounds (model,'EX_o2(e)',-1000,'l');
Change glucose uptake rate to 10 mmol gDW^{-1} h^{-1}	model = changeRxnBounds (model,'EX_glc(e)',-10,'l');
Simulate maximum growth by FBA	solution = optimizeCbModel (model);
Simulate maximum growth of regulated model	[FBAsols,DRgenes, constrainedRxns,cycleStart,states] = optimizeRegModel(model);
Change objective to maximum ATP yield	model = changeObjective (model,'ATPM');

using flux balance analysis (2). By constraining the glucose uptake rate to exactly −1 mmol gDW^{-1} h^{-1}, and setting the ATP maintenance reaction, *ATPM*, as the objective to be maximized, the yield of ATP from glucose can be calculated. *ATPM* is a stoichiometrically balanced reaction that drains ATP from the network. To determine the maximum yields of NADH and NADPH, similar balanced drain reactions must be added to the network and set as objectives. The maximum yields of these cofactors are given in Table 18. These yields are limited by the balancing of protons and by the stoichiometry of the network. See Table 19 for examples of COBRA Toolbox (157) commands for performing these simulations.

High-Throughput Data Analysis

A metabolic network can be used as a tool to analyze different types of high-throughput data, including gene expression data. These data can be mapped to the network, providing a context in which to interpret the results. In this example, gene expression data for *E. coli* comparing anaerobic growth with aerobic growth on glucose (58) were mapped to the core *E. coli* model (Fig. 20). In this figure, the triangles next to each reaction name represent the genes associated with each reaction. The GPRs in the core model were used to determine which reactions were upregulated or downregulated based on the gene regulation. When plotted against the network map, it is clear that glycolysis and the oxidative branch of the pentose phosphate pathway are upregulated, while the TCA cycle is downregulated. These results do not fully agree with the fluxes predicted by flux balance analysis in "Understanding Metabolic Capabilities," which is expected because gene expression and metabolic fluxes are not trivially quantitatively related. Nevertheless, such experimental data can be used to identify repressed genes and hence regulatory rules for a particular condition. Setting the flux to zero for reactions catalyzed by repressed genes allows for more biochemically realistic flux balance analysis.

A recently developed computational algorithm called GIMME (187) can be used to interpret gene expression data using constraint-based models. This algorithm uses data from microarrays to form context-specific models by adding reactions for which the associated gene expression levels are above a specified threshold. Additional reactions are then added to meet a known cellular function such as growth or secretion of a metabolite for the conditions in which the microarray data was gathered. These reactions are assigned an inconsistency score based on their agreement with the gene expression data. The GIMME algorithm and the *E. coli* core metabolic model were used along with 170 *E. coli* microarrays covering a large number of growth conditions and genetic perturbations to find which conditions are most consistent with the network topology needed for secretion of 12 different organic metabolites. This analysis reveals which conditions are most amenable to production of certain metabolites. Figure 21 shows the consistency scores for each context-specific model created. As we would expect, the common anaerobic secretion products such as ethanol, acetate, and formate (50) are more consistent with the arrays from anaerobically grown strains than with aerobic strains.

Discovery

Metabolic and regulatory models can be used to facilitate biological discovery. Computational predictions can be compared with experimental measurements under many different conditions, and when there are disagreements, it is because the model may be incomplete or incorrect in some way. These disagreements can be analyzed, aiding in the discovery of new biological features. Usually, the most up-to-date genome-scale models are used for discovery purposes. The core *E. coli* model is not suitable for discovery because it intentionally lacks most known metabolic reactions and has a limited scope.

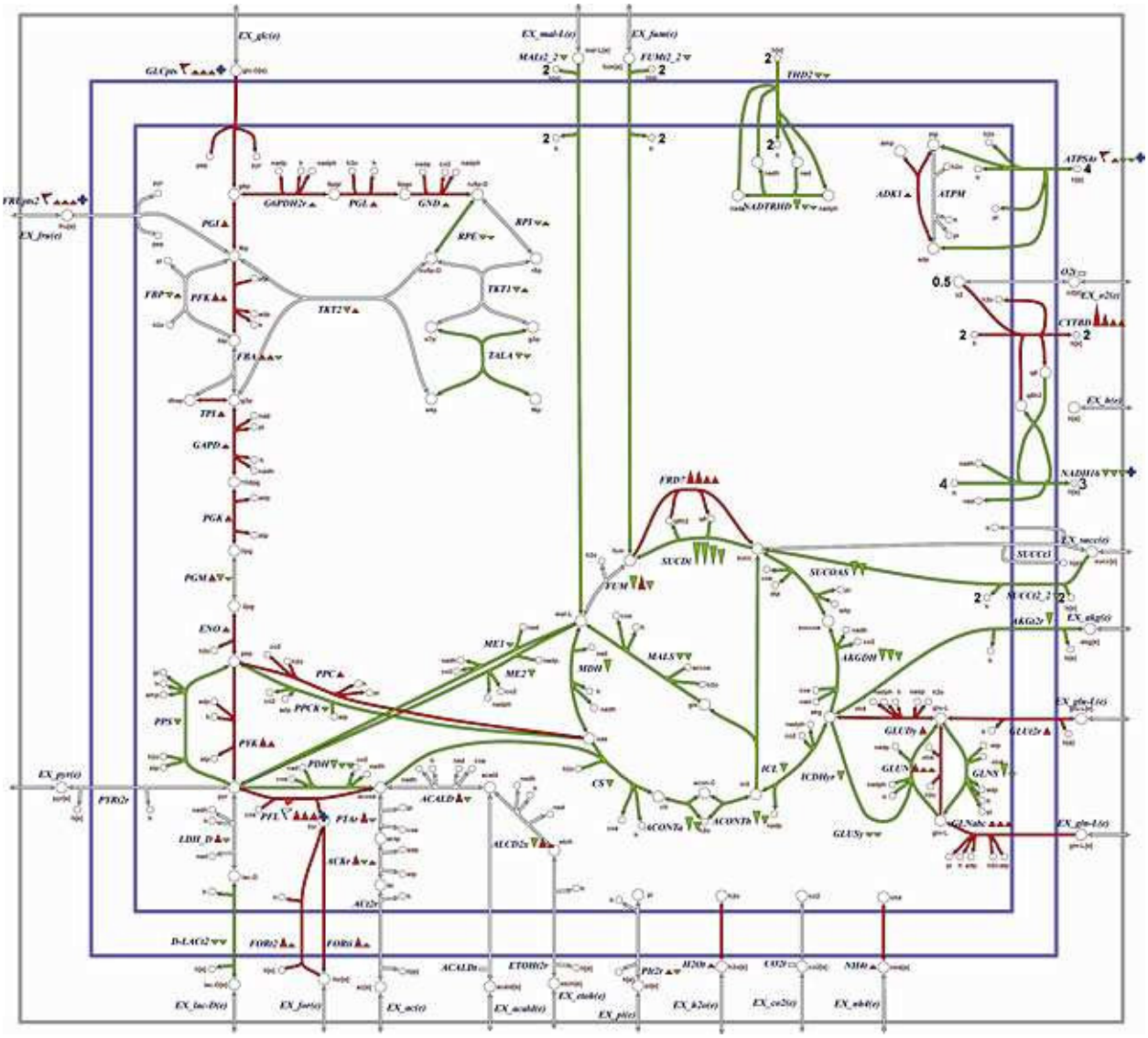

Figure 20 Gene expression data comparing anaerobic and aerobic growth mapped to the core *E. coli* model. Triangles next to reaction names represent the genes associated with each reaction. Red genes and reactions are upregulated under anaerobic conditions, and green genes and reactions are downregulated.

Models have been used most often to characterize unknown ORFs in an organism's genome. An algorithm that combines computational analysis of the genome-scale metabolic model *i*JR904 (188) with experimental growth phenotype screening was recently used to identify eight unknown ORFs in *E. coli* (22). First, *E. coli* was grown under many different minimal media conditions in a high-throughput screen with different carbon and nitrogen sources. The growth phenotypes were qualitatively compared with growth phenotypes predicted by flux balance analysis using the genome-scale *E. coli* metabolic model. The comparison highlighted 50 conditions where growth is possible in vivo but not possible in silico. For every growth phenotype that could not be explained by the model, an optimization algorithm was used to determine the minimum number of reactions that needed to be added to the model from a universal database of reactions (from KEGG [37]) to make growth possible.

In this optimization algorithm, the stoichiometric matrix, representing *E. coli* reactions and metabolites, was used in addition to a second matrix, containing all the reactions from the universal database, and a third matrix, containing exchange reactions for metabolites not included in

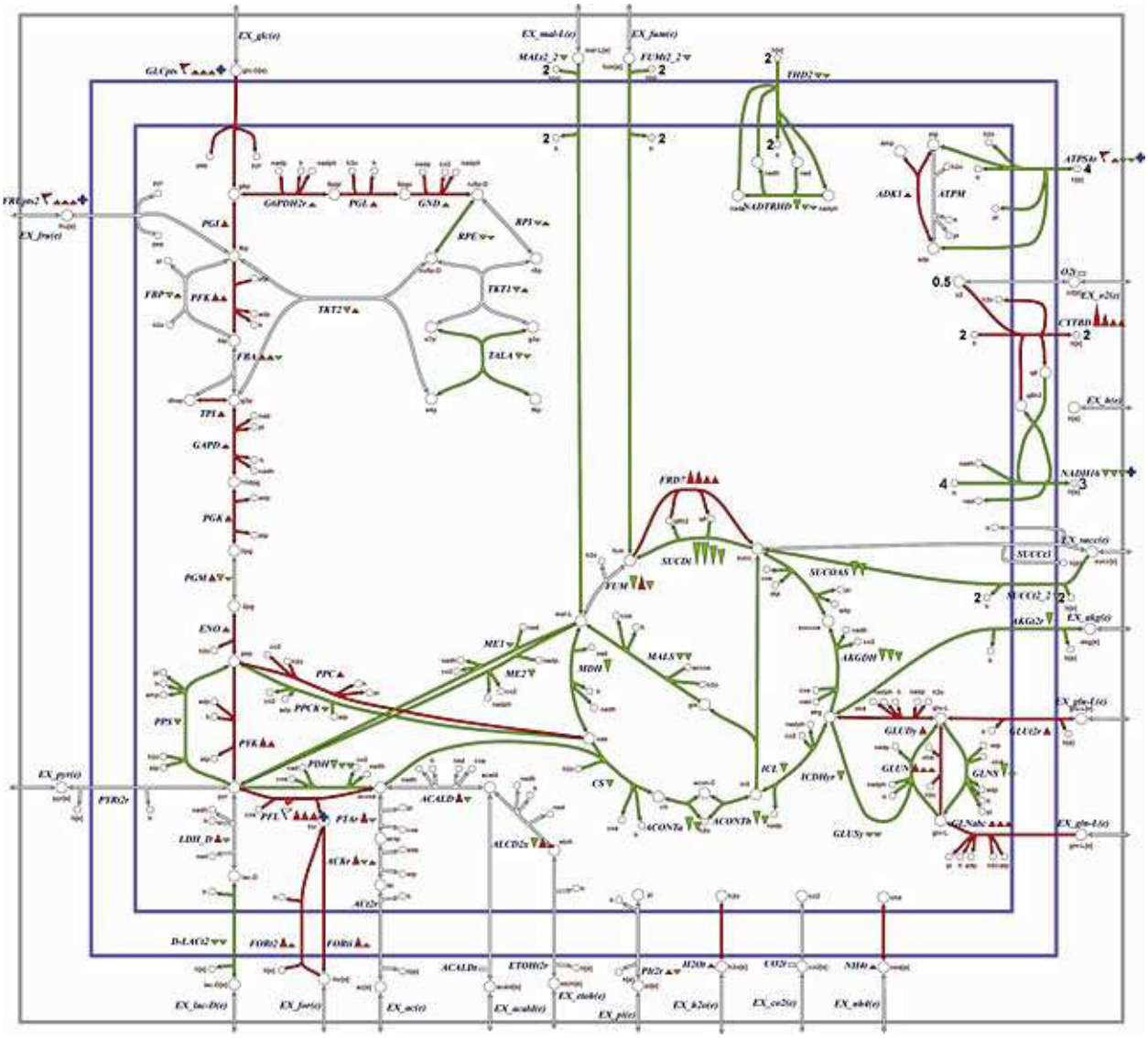

Figure 21 Normalized consistency scores for the secretion of 12 different products from the core *E. coli* model as computed from 170 microarrays (187). A high consistency score is given when the microarray data are in good agreement with the network topology necessary for the secretion of the product of interest. Each row has been normalized so that consistency scores within rows can be compared, but comparisons cannot be made between rows. Arrays 1 to 37 are wild type regulatory gene-deletion strains of *E. coli* grown on glucose under anaerobic conditions. Arrays 38 to 65 represent strains in which metabolic genes were knocked out followed by anaerobic growth on glucose. Arrays 66 to 77 are from regulatory gene-deletion mutants and wild-type strains grown anaerobically on glucose with nitrate used as the terminal electron acceptor. Arrays 78 to 98 and 148 to 170 are glucose aerobic wild-type and gene deletions of both regulatory proteins and metabolic enzymes. Arrays 99 to 147 are from wild-type strains that were under aerobically in lactate. Consistency scores are higher for production of most of the anaerobic fermentation products when using microarray data from anaerobic conditions, indicating that these conditions are more likely to lead to production of these products.

*i*JR904. For 26 of the 50 tested conditions, at least one set of universal database and/or exchange reactions were predicted to be necessary for in silico growth. In some growth conditions, up to 15 different sets of reactions were predicted to each allow in silico growth. These new reaction sets served as hypotheses for the identities of unknown ORFs, and several of these hypotheses were experimentally investigated in more detail.

The algorithm predicted that transport reactions needed to be added to the model to allow growth on propionate and 5-keto-D-gluconate. Eight genes that were predicted

to be transporters, by homology search, were tested experimentally. It was found that *E. coli* cannot grow on propionate without the gene *putP*, and cannot grow on 5-keto-D-gluconate without *idnT*, indicating that these genes do code for transporters. The algorithm also predicted that growth on D-malate is possible with a transport reaction and enzymatic conversion to succinate. *E. coli* strains lacking the genes *dctA*, *yeaT*, or *yeaU* are unable to grow on D-malate. Through analysis of gene expression data and biochemical assays, it was found that *dctA* is the transporter, *yeaU* is the enzyme, and *yeaT* is a regulator of the enzyme.

In this same study, genes involved in growth on galactonate γ-lactone were identified through gene expression data, even though the algorithm was unable to identify the necessary reactions. The genome-scale model was still useful, however, because it demonstrated that no known genes were responsible for growth on this substrate. Orphan reactions, reactions that are known to exist but are catalyzed by unknown enzymes, provide another way to identify unknown ORFs in microbial organisms. In the core *E. coli* model, there are four orphan transport reactions (listed in Table 20). Algorithms that consider the phylogenetic profiles (189) or coexpression and clustering on the genome (190) of the reactions adjacent the orphan reactions in the network have been shown to be successful in assigning tentative ORFs to such reactions.

Recently, an entirely new pathway for pyrimidine catabolism was discovered in *E. coli* by use of a combination of a subsystem approach to genome annotation and experimental validation (191). The subsystem approach to genome annotation and pathway analysis allows integrated knowledge of existing biochemical network structure and genomic structure to be projected across the entire collection of diverse species with completely sequenced genomes (192). In addition to establishing which organisms implement one or the other functional variants of a subsystem, this approach helps to reveal gaps in knowledge (missing genes) and potential new players (predicted genes) (193).

Table 20 Orphan reactions in the core *E. coli* reconstruction[a]

Abbr.	Full name	Equation
ACt2r	Acetate reversible transport via proton symport	ac[e] + h[e] ⟷ ac + h
ETOHt2r	Ethanol reversible transport via proton symport	etoh[e] + h[e] ⟷ etoh + h
PYRt2r	Pyruvate reversible transport via proton symport	h[e] + pyr[e] ⟷ h + pyr
SUCCt3	Succinate transport out via proton antiport	h[e] + succ ⟷ h + succ[e]

[a] Orphan reactions are reactions that are known to exist but that are catalyzed by enzymes without an assigned ORF.

Synthetic Biology and Metabolic Engineering

Because of the predictive capabilities of constraint-based metabolic models, these models can be used in synthetic biology and metabolic engineering applications (158, 194). In particular, constraint-based models are useful because they can predict when a particular metabolite will be overproduced and secreted. Gene knockouts and knock-ins can be simulated by removing or adding reactions to the network, and the behavior of the modified network can be predicted by flux balance analysis. These model-based predictions are more accurate than predictions based in intuition and knowledge of gene functions, because flux balance analysis considers the complicated interacting effects that a knockout has on all pathways simultaneously.

Growth-coupled designs are a particularly promising class of metabolically engineered strains. When the product of interest in one of these strains is produced at a higher rate, the growth rate of the strain also increases. This is unlike most strain designs, in which the growth rate decreases as flux is diverted to a metabolic by-product. Growth-coupled designs are evolutionarily stable, because their production rates actually increase as mutations that increase growth rate accumulate. The growth coupling of designs can be visualized by using a production envelope, a graph that shows the solution space of a model in the dimensions of growth rate and the exchange reaction of a particular metabolite. When the metabolite is growth coupled, its minimum exchange rate at the maximum possible growth rate will be greater than zero. OptKnock is a bilevel linear programming optimization algorithm that uses a constraint-based model to identify sets of reaction knockouts that couple the production of a metabolite to growth (27). It seeks to simultaneously maximize growth rate and product secretion rate. This algorithm was used to identify growth-coupled designs for many metabolites in the core *E. coli* model under anaerobic conditions. The production envelopes of some growth-coupled designs are shown in Fig. 22.

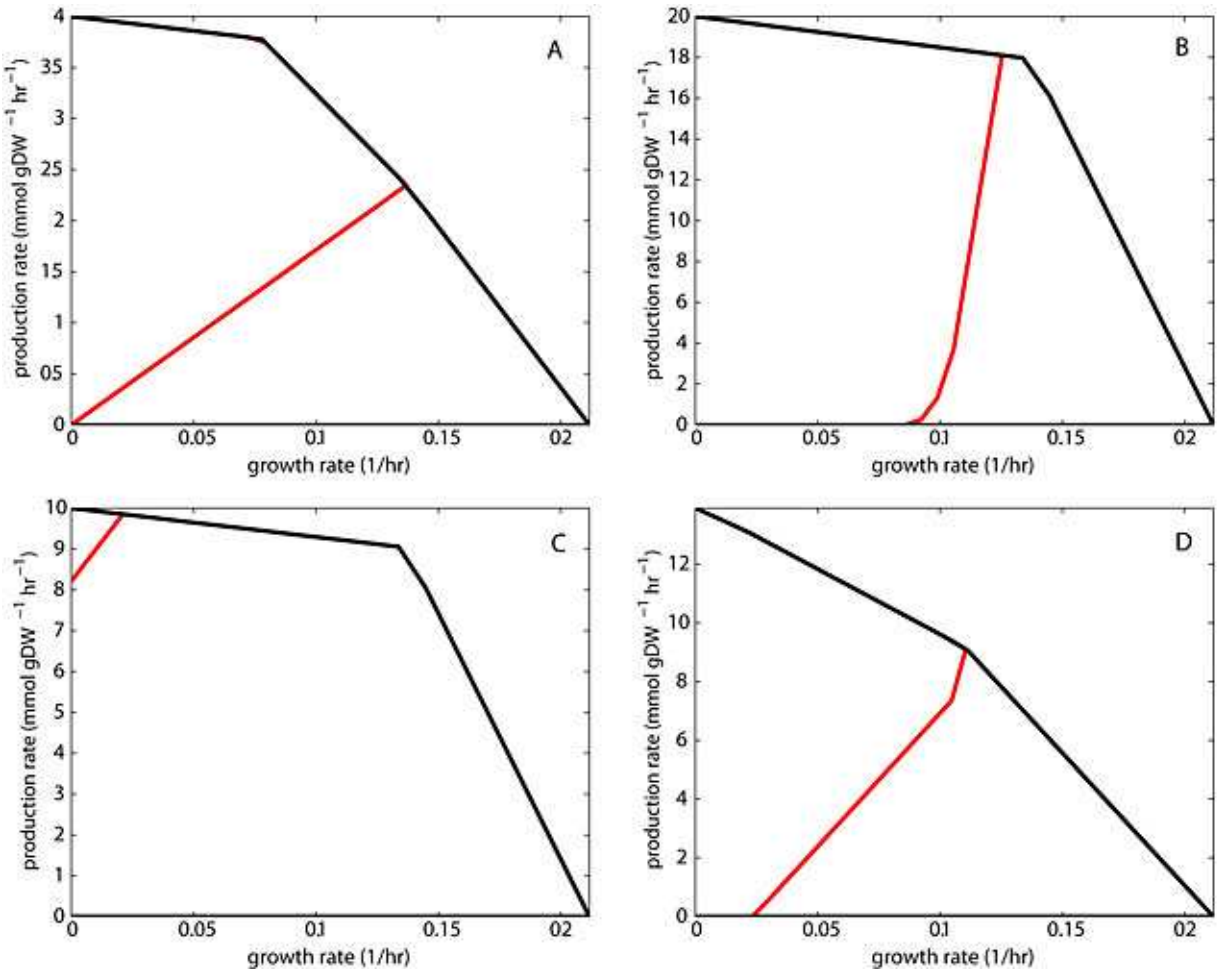

Figure 22 Production envelopes for OptKnock designs for the production of 2-oxoglutarate (A), D-lactate (B), pyruvate (C), and succinate (D). Black envelopes are wild type, and red envelopes are for knockout strains. Growth is possible at any point inside the production envelope, but over time *E. coli* will adapt to the highest possible growth rate (26).

SUMMARY

Currently, the wealth of biochemical information exceeds the scope and depth of even the largest network reconstructions. The scope of the *E. coli* core model represents perhaps the most well characterized fraction of the latest *E. coli* genome-scale metabolic model (4). Nevertheless, the results of future genetic studies that establish the genes corresponding to orphan reactions need to be incorporated as part of an iterative cycle of development (158). The regulatory network for *E. coli* core metabolism is rapidly being discovered by use of high-throughput protein-DNA-binding assays (64, 195, 196, 197). The incorporation of this new knowledge, with a more comprehensive study of the biochemical literature on core metabolic regulation, would result in a significant expansion in the number and complexity of Boolean regulatory rules in the *E. coli* core model.

Biochemical network reconstructions are biochemically, genetically, and genomically structured databases. These reconstructions rely heavily on information retrieved from biochemical characterization of reactions and their substrates. Genetic studies are essential to identify the ORF(s) that encode the enzyme(s) responsible for each catalytic reaction, giving rise to GPRs. More broadly, species-specific biochemical and molecular biological studies are essential to provide a sound experimental basis for the components and reactions that are the key elements in any reconstruction. The utility of biochemical network reconstructions is driven by the fact that they can be transformed into a computational model. In turn, the computational model can be applied to address an increasingly wide variety of biological questions (158), including bacterial evolution (11), analysis of network properties (13, 12, 15), study of phenotypic behavior (17,

20), biological discovery (22, 23, 58), and metabolic engineering (25, 27, 28).

In the future, we can look forward to complete reconstructions of all known biochemical processes in *E. coli* and *Salmonella*. Such models ideally serve as structured self-consistent representations of our knowledge. Inevitably they will grow in scope, and more details will need to be added such that model predictions finally reach par with experimental observations. Even now, in certain situations, model predictions not only provide qualitatively and quantitatively accurate predictions of experiments, but can also be used to suggest profitable avenues for experimental confirmation. Even though reconstructions continue to grow in size and scope, fundamentally, they will still be systems of biochemical reactions, just like the core *E. coli* model, but with the rest of the cell built around it. Ultimately, the full utility of computational models will be realized at the fingertips of biological domain-specific experts. However, if all other domain-specific experts have contributed to the same model, as each expert probes her or his own area of biological expertise in silico, they can do so in the knowledge that their in silico predictions automatically account for others' expertise, through the biochemically, genetically, and genomically structured model. We hope that this introduction to the core *E. coli* model brings this era a little closer.

SUPPLEMENTAL FILES

The following supplemental files can be found online at http://systemsbiology.ucsd.edu/:

- **ecoli_core_model.xls** A Microsoft Excel file that describes the reactions and metabolites in the core *E. coli* model. This file includes the full S matrix.
- **core_regulatory_rules.xls** A Microsoft Excel file that gives the Boolean regulatory rules for every gene in the regulated model.
- **ecoli_core_model.mat** A MATLAB data file that contains the core *E. coli* model (without regulation) in a format that can be used with the COBRA Toolbox (see Becker et al. [157] for details on this toolbox).
- **modelReg.mat** A MATLAB data file that contains the regulated core *E. coli* model for the COBRA Toolbox.
- **optimizeRegModel.m** A new COBRA Toolbox function that is needed to run simulations using the regulated model. This function uses flux balance analysis to determine the state of the network while considering regulatory constraints.
- **dynamicRFBA.m** A new COBRA Toolbox function that performs dynamic flux balance analysis using the regulated model (see Covert et al. [69] for details on dynamic rFBA).
- **solveBooleanRegModel.m** A COBRA Toolbox function that is required by both optimizeRegModel.m and dynamicRFBA.m.
- **ecoli_core_model.xml** An SBML file for the core *E. coli* model without regulation. COBRA Toolbox users with the SBML Toolbox can load this file using the function "readCbModel."

ACKNOWLEDGMENTS

We thank B. K. Cho, N. Lewis, I. Thiele, and K. Zengler for their helpful comments.

No potential conflicts of interest relevant to this review were reported.

REFERENCES

1. Palsson BØ. 2006. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, New York, NY.

2. Varma A, Palsson BØ. 1993. Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *J Theor Biol* **165:**477–502.

3. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ. 2007. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* **104:**1777–1782.

4. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3:**121.

5. Nogales J, Palsson BØ, Thiele I. 2008. A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory. *BMC Syst Biol* **2:**79.

6. Resendis-Antonio O, Reed JL, Encarnacion S, Collado-Vides J, Palsson BØ. 2007. Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. *PLoS Comput. Biol.* **3:**1887–1895.

7. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. 2009. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7:**129–143.

8. Price ND, Papin JA, Schilling CH, Palsson B. 2003. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* **21:**162–169.

9. Pal C, Papp B, Lercher MJ. 2005. Horizontal gene transfer depends on gene content of the host. *Bioinformatics* **21**(Suppl. 2):ii222–ii223.

10. Pal C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* **37:**1372–1375.

11. Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. 2006. Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440:**667–670.

12. Mahadevan R, Schilling CH. 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* **5:**264–276.

13. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* **14:**301–312.

14. Barrett CL, Herring CD, Reed JL, Palsson BO. 2005. The global transcriptional regulatory network for metabolism in *Escherichia coli* attains few dominant functional states. *Proc Natl Acad Sci USA* **102:**19103–19108.

15. Samal A, Jain S. 2008. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. *BMC Syst Biol* **2:**21.

16. Almaas E, Kovacs B, Vicsek T, Oltvai ZN, Barabasi AL. 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427:**839–843.

17. Edwards JS, Ibarra RU, Palsson BØ. 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* **19:**125–130.

18. Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* **99:**15112–15117.

19. Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* **102:**7695–7700.

20. Ibarra RU, Edwards JS, Palsson BØ. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420:**186–189.

21. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BO, Agarwalla S. 2006. Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J Bacteriol* **188:**8259–8271.

22. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BØ. 2006. Systems approach to refining genome annotation. *Proc Natl Acad Sci USA* **103:**17480–17484.

23. Herrgard MJ, Fong SS, Palsson BØ. 2006. Identification of genome-scale metabolic network models using experimentally measured flux profiles. *PLoS Comput Biol* **2:**e72.

24. Kumar VS, Maranas CD. 2009. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput Biol* **5:**e1000308.

25. Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proc Natl Acad Sci USA* **104:**7797–7802.

26. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, Maranas CD, Palsson BO. 2005. *In silico* design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnol Bioeng* **91:**643–648.

27. Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng* **84:**647–657.

28. Patil KR, Rocha I, Forster J, Nielsen J. 2005. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics* **6:**308.

29. Karp PD, Keseler IM, Shearer A, Latendresse M, Krummenacker M, Paley SM, Paulsen I, Collado-Vides J, Gama-Castro S, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Bonavides-Martinez C, Ingraham J. 2007. Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res* **35:**7577–7590.

30. Stein L. 2001. Genome annotation: from sequence to biology. *Nat Rev Genet* **2:**493–503.

31. Christie KR, Weng S, Balakrishnan R, Costanzo MC, Dolinski K, Dwight SS, Engel SR, Feierbach B, Fisk DG, Hirschman JE, Hong EL, Issel-Tarver L, Nash R, Sethuraman A, Starr B, Theesfeld CL, Andrada R, Binkley G, Dong Q, Lane C, Schroeder M, Botstein D, Cherry JM. 2004. Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms. *Nucleic Acids Res* **32**(Database issue)**:**D311–D314.

32. Guldener U, Munsterkotter M, Kastenmuller G, Strack N, van Helden J, Lemer C, Richelles J, Wodak SJ, Garcia-Martinez J, Perez-Ortin JE, Michael H, Kaps A, Talla E, Dujon B, Andre B, Souciet JL, De Montigny J, Bon E, Gaillardin C, Mewes HW. 2005. CYGD: the Comprehensive Yeast Genome Database. *Nucleic Acids Res* **33:**D364–D368.

33. Maglott D, Ostell J, Pruitt KD, Tatusova T. 2005. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* **33:**D54–D58.

34. Peterson JD, Umayam LA, Dickinson T, Hickey EK, White O. 2001. The comprehensive microbial resource. *Nucleic Acids Res* **29:**123–125.

35. Stoesser G, Tuli MA, Lopez R, Sterk P. 1999. The EMBL nucleotide sequence database. *Nucleic Acids Res* **27:**18–24.

36. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I, Lykidis A, Mavromatis K, Ivanova N, Kyrpides NC. 2006. The integrated microbial genomes (IMG) system. *Nucleic Acids Res* **34:**D344–D348.

37. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28:**27–30.

38. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res* **37:**D588–D592.

39. Bairoch A. 2000. The ENZYME database in 2000. *Nucleic Acids Res* **28:**304–305.

40. Krieger CJ, Zhang P, Mueller LA, Wang A, Paley S, Arnaud M, Pick J, Rhee SY, Karp PD. 2004. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* **32**(Database issue):D438–D442.

41. DeJongh M, Formsma K, Boillot P, Gould J, Rycenga M, Best A. 2007. Toward the automated generation of genome-scale metabolic networks in the SEED. *BMC Bioinformatics* **8:**139.

42. Ren Q, Chen K, Paulsen IT. 2007. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res* **35:**D274–D279.

43. Paley SM, Karp PD. 2002. Evaluation of computational metabolic-pathway predictions for *Helicobacter pylori*. *Bioinformatics* **18:**715–724.

44. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. 2003. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res* **31:**6633–6639.

45. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:**25–29.

46. Reed JL, Famili I, Thiele I, Palsson BØ. 2006. Towards multidimensional genome annotation. *Nat Rev Genet* **7:**130–141.

47. Neidhardt FC, Curtis R III, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, Umbarger HE (ed). 1996. *Escherichia coli and Salmonella*: *Cellular and Molecular Biology*, 2nd ed. ASM Press, Washington, DC.

48. Edwards JS, Palsson BO. 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97:**5528–5533.

49. Varma A, Palsson BØ. 1993. Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. *J Theor Biol* **165:**503–522.

50. Varma A, Palsson BØ. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol* **60:**3724–3731.

51. Duarte NC, Herrgard MJ, Palsson B. 2004. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res* **14:**1298–1309.

52. Breitling R, Vitkup D, Barrett MP. 2008. New surveyor tools for charting microbial metabolic maps. *Nat Rev Microbiol* **6:**156–161.

53. Froman BE, Tait RC, Gottlieb LD. 1989. Isolation and characterization of the phosphoglucose isomerase gene from *Escherichia coli*. *Mol Gen Genet* **217:**126–131.

54. Hua Q, Yang C, Baba T, Mori H, Shimizu K. 2003. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J Bacteriol* **185:**7053–7067.

55. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, Longabaugh W, Vuthoori M, Whitehead K, Madar A, Suzuki L, Mori T, Chang DE, Diruggiero J, Johnson CH, Hood L, Baliga NS. 2007. A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131:**1354–1365.

56. Herrgard MJ, Lee BS, Portnoy V, Palsson BØ. 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res* **16:**627–635.

57. Workman CT, Mak HC, McCuine S, Tagne JB, Agarwal M, Ozier O, Begley TJ, Samson LD, Ideker T. 2006. A systems approach to mapping DNA damage response pathways. *Science* **312:**1054–1059.

58. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BØ. 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429:**92–96.

59. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* **5:**e8.

60. Hu Z, Killion PJ, Iyer VR. 2007. Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39:**683–687.

61. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N. 2003. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* **34:**166–176.

62. Cho BK, Barrett CL, Knight EM, Park YS, Palsson BO. 2008. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proc Natl Acad Sci USA* **105:**19462–19467.

63. Cho BK, Knight EM, Barrett CL, Palsson BØ. 2008. Genome-wide analysis of Fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. *Genome Res* **18:**900–910.

64. Cho BK, Knight EM, Palsson BØ. 2008. Genomewide identification of protein binding locations using chromatin immunoprecipitation coupled with microarray. *Methods Mol Biol* **439:**131–145.

65. Ideker TE, Thorsson V, Karp RM. 2000. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pacific Symp Biocomput* **292:**305–316.

66. Cho BK, Knight EM, Palsson BØ. 2006. Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA. *Microbiology* **152:**2207–2219.

67. Grainger DC, Aiba H, Hurd D, Browning DF, Busby SJ. 2007. Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucleic Acids Res* **35:**269–278.

68. Shimada T, Ishihama A, Busby SJ, Grainger DC. 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* **36:**3950–3955.

69. Covert MW, Schilling CH, Palsson B. 2001. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* **213:**73–88.

70. Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BØ. 2006. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* **2:**e101.

71. Gianchandani EP, Joyce AR, Palsson BO, Papin JA. 2009. Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. *PLoS Comput Biol* **5:**e1000403.

72. Thiele I, Jamshidi N, Fleming RM, Palsson BO. 2009. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol* **5:**e1000312.

73. Alberty RA. 2003. *Thermodynamics of Biochemical Reactions*. Massachusetts Institute of Technology, Cambridge, MA.

74. Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G III, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL. 2006. *Escherichia coli* K-12: a cooperatively developed annotation snapshot-2005. *Nucleic Acids Res* **34:**1–9.

75. Cunningham L, Guest JR. 1998. Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology* **144**(Pt 8):2113–2123.

76. Keseler IM, Collado-Vides J, Gama-Castro S, Ingraham J, Paley S, Paulsen IT, Peralta-Gil M, Karp PD. 2005. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res* **33:**D334–D337.

77. Stolz B, Huber M, Markovic-Housley Z, Erni B. 1993. The mannose transporter of *Escherichia coli*. Structure and function of the IIABMan subunit. *J Biol Chem* **268:**27094–27099.

78. Reidl J, Boos W. 1991. The *malX malY* operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an enzyme abolishing the endogenous induction of the maltose system. *J Bacteriol* **173:**4862–4876.

79. Eberstadt M, Grdadolnik SG, Gemmecker G, Kessler H, Buhr A, Erni B. 1996. Solution structure of the IIB domain of the glucose transporter of *Escherichia coli*. *Biochemistry* **35:**11286–11292.

80. Daldal F. 1984. Nucleotide sequence of gene *pfkB* encoding the minor phosphofructokinase of *Escherichia coli* K-12. *Gene* **28:**337–342.

81. Rypniewski WR, Evans PR. 1989. Crystal structure of unliganded phosphofructokinase from *Escherichia coli*. *J Mol Biol* **207:**805–821.

82. Hines JK, Fromm HJ, Honzatko RB. 2006. Novel allosteric activation site in *Escherichia coli* fructose-1,6-bisphosphatase. *J Biol Chem* **281:**18386–18393.

83. Alefounder PR, Baldwin SA, Perham RN, Short NJ. 1989. Cloning, sequence analysis and over-expression of the gene for the class II fructose 1,6-bisphosphate aldolase of *Escherichia coli*. *Biochem J* **257:**529–534.

84. Baldwin SA, Perham RN. 1978. Novel kinetic and structural properties of the class-I D-fructose 1,6-bisphosphate aldolase from *Escherichia coli* (Crookes' strain). *Biochem J* **169:**643–652.

85. Thomson GJ, Howlett GJ, Ashcroft AE, Berry A. 1998. The *dhnA* gene of *Escherichia coli* encodes a class I fructose bisphosphate aldolase. *Biochem J* **331**(Pt. 2)**:**437–445.

86. Pichersky E, Gottlieb LD, Hess JF. 1984. Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*. *Mol Gen Genet* **195:**314–320.

87. Branlant G, Branlant C. 1985. Nucleotide sequence of the *Escherichia coli gap* gene. Different evolutionary behavior of the NAD^+-binding domain and of the catalytic domain of D-glyceraldehyde-3-phosphate dehydrogenase. *Eur J Biochem* **150:**61–66.

88. Nellemann LJ, Holm F, Atlung T, Hansen FG. 1989. Cloning and characterization of the *Escherichia coli* phosphoglycerate kinase (*pgk*) gene. *Gene* **77:**185–191.

89. Fraser HI, Kvaratskhelia M, White MF. 1999. The two analogous phosphoglycerate mutases of *Escherichia coli*. *FEBS Lett* **455:** 344–348.

90. Kuhnel K, Luisi BF. 2001. Crystal structure of the *Escherichia coli* RNA degradosome component enolase. *J Mol Biol* **313:**583–592.

91. Garrido-Pertierra A, Cooper RA. 1983. Evidence for two distinct pyruvate kinase genes in *Escherichia coli* K-12. *FEBS Lett* **162:**420–422.

92. Muirhead H. 1990. Isoenzymes of pyruvate kinase. *Biochem Soc Trans* **18:**193–196.

93. Josephson BL, Fraenkel DG. 1974. Sugar metabolism in transketolase mutants of *Escherichia coli*. *J Bacteriol* **118:**1082–1089.

94. Josephson BL, Fraenkel DG. 1969. Transketolase mutants of *Escherichia coli*. *J Bacteriol* **100:**1289–1295.

95. Peyru G, Fraenkel DG. 1968. Genetic mapping of loci for glucose-6-phosphate dehydrogenase, gluconate-6-phosphate dehydrogenase, and gluconate-6-phosphate dehydrase in *Escherichia coli*. *J Bacteriol* **95:**1272–1278.

96. Thomason LC, Court DL, Datta AR, Khanna R, Rosner JL. 2004. Identification of the *Escherichia coli* K-12 *ybhE* gene as *pgl*, encoding 6-phosphogluconolactonase. *J Bacteriol* **186:**8248–8253.

97. Veronese FM, Boccu E, Fontana A. 1976. Isolation and properties of 6-phosphogluconate dehydrogenase from *Escherichia coli*. Some comparisons with the thermophilic enzyme from *Bacillus stearothermophilus*. *Biochemistry* **15:**4026–4033.

98. Essenberg MK, Cooper RA. 1975. Two ribose-5-phosphate isomerases from *Escherichia coli* K12: partial characterisation of the enzymes and consideration of their possible physiological roles. *Eur J Biochem* **55:**323–332.

99. Csonka LN, Fraenkel DG. 1977. Pathways of NADPH formation in *Escherichia coli*. *J Biol Chem* **252:**3382–3391.

100. Melendez-Hevia E, Isidoro A. 1985. The game of the pentose phosphate cycle. *J Theor Biol* **117:**251–263.

101. Iida A, Teshiba S, Mizobuchi K. 1993. Identification and characterization of the *tktB* gene encoding a second transketolase in *Escherichia coli* K-12. *J Bacteriol* **175:**5375–5383.

102. Sprenger GA, Schorken U, Sprenger G, Sahm H. 1995. Transketolase A of *Escherichia coli* K12. Purification and properties of the enzyme from recombinant strains. *Eur J Biochem* **230:**525–532.

103. Lyngstadaas A, Sprenger GA, Boye E. 1998. Impaired growth of an *Escherichia coli rpe* mutant lacking ribulose-5-phosphate epimerase activity. *Biochim Biophys Acta* **1381:**319–330.

104. Reed LJ, Pettit FH, Eley MH, Hamilton L, Collins JH, Oliver RM. 1975. Reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex. *Proc Natl Acad Sci USA* **72:**3068–3072.

105. Nguyen NT, Maurus R, Stokell DJ, Ayed A, Duckworth HW, Brayer GD. 2001. Comparative analysis of folding and substrate binding sites between regulated hexameric type II citrate synthases and unregulated dimeric type I enzymes. *Biochemistry* **40:**13177–13187.

106. Brock M, Maerker C, Schutz A, Volker U, Buckel W. 2002. Oxidation of propionate to pyruvate in *Escherichia coli*. Involvement of methylcitrate dehydratase and aconitase. *Eur J Biochem* **269:**6184–6194.

107. Prodromou C, Haynes MJ, Guest JR. 1991. The aconitase of *Escherichia coli*: purification of the enzyme and molecular cloning and map location of the gene (*acn*). *J Gen. Microbiol.* **137:**2505–2515.

108. Burke WF, Johanson RA, Reeves HC. 1974. $NADP^+$-specific isocitrate dehydrogenase of *Escherichia coli*. II. Subunit structure. *Biochim Biophys Acta* **351:**333–340.

109. Perham RN, Packman LC. 1989. 2-Oxo acid dehydrogenase multienzyme complexes: domains, dynamics, and design. *Ann N Y Acad Sci* **573:**1–20.

110. Bridger WA, Wolodko WT, Henning W, Upton C, Majumdar R, Williams SP. 1987. The subunits of succinyl-coenzyme A synthetase—function and assembly. *Biochem Soc Symp* **54:**103–111.

111. Yankovskaya V, Horsefield R, Tornroth S, Luna-Chavez C, Miyoshi H, Leger C, Byrne B, Cecchini G, Iwata S. 2003. Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299:**700–704.

112. Condon C, Cammack R, Patil DS, Owen P. 1985. The succinate dehydrogenase of *Escherichia coli*. Immunochemical resolution and biophysical characterization of a 4-subunit enzyme complex. *J Biol Chem* **260:**9427–9434.

113. Bell PJ, Andrews SC, Sivak MN, Guest JR. 1989. Nucleotide sequence of the FNR-regulated fumarase gene (*fumB*) of *Escherichia coli* K-12. *J Bacteriol* **171:**3494–3503.

114. Flint DH. 1994. Initial kinetic and mechanistic characterization of *Escherichia coli* fumarase A. *Arch Biochem Biophys* **311:**509–516.

115. Woods SA, Schwartzbach SD, Guest JR. 1988. Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochim Biophys Acta* **954:**14–26.

116. Sutherland P, McAlister-Henn L. 1985. Isolation and expression of the *Escherichia coli* gene encoding malate dehydrogenase. *J Bacteriol* **163:**1074–1079.

117. Cole ST, Condon C, Lemire BD, Weiner JH. 1985. Molecular biology, biochemistry and bioenergetics of fumarate reductase, a complex membrane-bound iron-sulfur flavoenzyme of *Escherichia coli*. *Biochim Biophys Acta* **811:**381–403.

118. Iverson TM, Luna-Chavez C, Cecchini G, Rees DC. 1999. Structure of the *Escherichia coli* fumarate reductase respiratory complex. *Science* **284:**1961–1966.

119. Hoyt JC, Robertson EF, Berlyn KA, Reeves HC. 1988. *Escherichia coli* isocitrate lyase: properties and comparisons. *Biochim Biophys Acta* **966:**30–35.

120. Molina I, Pellicer MT, Badia J, Aguilar J, Baldoma L. 1994. Molecular characterization of *Escherichia coli* malate synthase G.

Differentiation with the malate synthase A isoenzyme. *Eur J Biochem* **224:**541–548.

121. Narindrasorasak S, Bridger WA. 1977. Phosphoenolpyruvate synthetase of *Escherichia coli*: molecular weight, subunit composition, and identification of phosphohistidine in phosphoenzyme intermediate. *J Biol Chem* **252:**3121–3127.

122. Cooper RA, Kornberg HL. 1965. Net formation of phosphoenolpyruvate from pyruvate by *Escherichia coli*. *Biochim Biophys Acta* **104:**618–620.

123. Hansen EJ, Juni E. 1975. Isolation of mutants of *Escherichia coli* lacking NAD- and NADP-linked malic. *Biochem Biophys Res Commun* **65:**559–566.

124. Hansen EJ, Juni E. 1974. Two routes for synthesis of phosphoenolpyruvate from C4-dicarboxylic acids in *Escherichia coli*. *Biochem Biophys Res Commun* **59:**1204–1210.

125. Iwakura M, Hattori J, Arita Y, Tokushige M, Katsuki H. 1979. Studies on regulatory functions of malic enzymes. VI. Purification and molecular properties of NADP-linked malic enzyme from *Escherichia coli* W. *J Biochem* **85:**1355–1365.

126. Mahajan SK, Chu CC, Willis DK, Templin A, Clark AJ. 1990. Physical analysis of spontaneous and mutagen-induced mutants of *Escherichia coli* K-12 expressing DNA exonuclease VIII activity. *Genetics* **125:**261–273.

127. Niersbach M, Kreuzaler F, Geerse RH, Postma PW, Hirsch HJ. 1992. Cloning and nucleotide sequence of the *Escherichia coli* K-12 ppsA gene, encoding PEP synthase. *Mol Gen Genet* **231:**332–336.

128. Delbaere LT, Sudom AM, Prasad L, Leduc Y, Goldie H. 2004. Structure/function studies of phosphoryl transfer by phosphoenolpyruvate carboxykinase. *Biochim Biophys Acta* **1697:**271–278.

129. Kai Y, Matsumura H, Izui K. 2003. Phosphoenolpyruvate carboxylase: three-dimensional structure and molecular mechanisms. *Arch Biochem Biophys* **414:**170–179.

130. Kornberg HL. 1966. Anaplerotic sequences and their role in metabolism. *Essays Biochem* **2:**1–31.

131. Kornberg HL. 1965. The coordination of metabolic routes. *Function and Structure in Microorganisms: Fifteenth Symposium of the Society for General Microbiology*. University Press, London, United Kingdom.

132. Wilks JC, Slonczewski JL. 2007. pH of the cytoplasm and periplasm of *Escherichia coli*: rapid measurement by green fluorescent protein fluorimetry. *J Bacteriol* **189:**5601–5607.

133. Unden G, Dunnwald P. 11 March 2008, posting date. The Aerobic and Anaerobic Respiratory Chain of Escherichia coli and Salmonella enterica: Enzymes and Energetics, The Aerobic and Anaerobic Respiratory Chain of *Escherichia coli* and *Salmonella enterica*: Enzymes and Energetics. *In* Böck A, Curtiss R III, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL, and Ussery D (ed), *EcoSal—Escherichia coli and Salmonella*: *Cellular and Molecular Biology*. ASM Press, Washington, DC.

134. Schneider D, Pohl T, Walter J, Dorner K, Kohlstadt M, Berger A, Spehr V, Friedrich T. 2008. Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I). *Biochim Biophys Acta* **1777:**735–739.

135. Spehr V, Schlitt A, Scheide D, Guenebaut V, Friedrich T. 1999. Overexpression of the *Escherichia coli* nuo-operon and isolation of the overproduced NADH:ubiquinone oxidoreductase (complex I). *Biochemistry* **38:**16261–16267.

136. Kobayashi K, Tagawa S, Mogi T. 1999. Electron transfer process in cytochrome bd-type ubiquinol oxidase from *Escherichia coli* revealed by pulse radiolysis. *Biochemistry* **38:**5913–5917.

137. Cain BD, Simoni RD. 1989. Proton translocation by the F1F0 ATPase of *Escherichia coli*. Mutagenic analysis of the a subunit. *J Biol Chem* **264:**3292–3300.

138. Kasimoglu E, Park SJ, Malek J, Tseng CP, Gunsalus RP. 1996. Transcriptional regulation of the proton-translocating ATPase (*atpIBEFHAGDC*) operon of *Escherichia coli*: control by cell growth rate. *J Bacteriol* **178:**5563–5567.

139. Berry MB, Bae E, Bilderback TR, Glaser M, Phillips GN Jr. 2006. Crystal structure of ADP/AMP complex of *Escherichia coli* adenylate kinase. *Proteins* **62:**555–556.

140. Brune M, Schumann R, Wittinghofer F. 1985. Cloning and sequencing of the adenylate kinase gene (*adk*) of *Escherichia coli*. *Nucleic Acids Res* **13:**7139–7151.

141. Bizouarn T, Fjellstrom O, Meuller J, Axelsson M, Bergkvist A, Johansson C, Goran Karlsson B, Rydstrom J. 2000. Proton translocating nicotinamide nucleotide transhydrogenase from *E. coli*. Mechanism of action deduced from its structural and catalytic properties. *Biochim Biophys Acta* **1457:**211–228.

142. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E. 2004. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* **279:**6613–6619.

143. Clark DP. 1989. The fermentation pathways of *Escherichia coli*. *FEMS Microbiol Rev* **5:**223–234.

144. Dym O, Pratt EA, Ho C, Eisenberg D. 2000. The crystal structure of D-lactate dehydrogenase, a peripheral membrane respiratory enzyme. *Proc Natl Acad Sci USA* **97:**9413–9418.

145. Jiang GR, Nikolova S, Clark DP. 2001. Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology* **147:**2437–2446.

146. Knappe J, Sawers G. 1990. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol Rev* **6:**383–398.

147. Sawers G, Watson G. 1998. A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. *Mol Microbiol* **29:**945–954.

148. Suzuki T. 1969. Phosphotransacetylase of *Escherichia coli* B, activation by pyruvate and inhibition by NADH and certain nucleotides. *Biochim Biophys Acta* **191:**559–569.

149. Skarstedt MT, Silverstein E. 1976. *Escherichia coli* acetate kinase mechanism studied by net initial rate, equilibrium, and independent isotopic exchange kinetics. *J Biol Chem* **251:**6775–6783.

150. Ferrandez A, Garcia JL, Diaz E. 1997. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl) propionate catabolic pathway of *Escherichia coli* K-12. *J Bacteriol* **179:** 2573–2581.

151. Kessler D, Herth W, Knappe J. 1992. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. *J Biol Chem* **267:**18073–18079.

152. Yuan J, Fowler WU, Kimball E, Lu W, Rabinowitz JD. 2006. Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. *Nat Chem Biol* **2:**529–530.

153. Wu LF, Mandrand-Berthelot MA. 1995. A family of homologous substrate-binding proteins with a broad range of substrate specificity and dissimilar biological functions. *Biochimie* **77:**744–750.

154. Wallace B, Yang YJ, Hong JS, Lum D. 1990. Cloning and sequencing of a gene encoding a glutamate and aspartate carrier of *Escherichia coli* K-12. *J Bacteriol* **172:**3214–3220.

155. Rhee SG, Ubom GA, Hunt JB, Chock PB. 1982. Catalytic cycle of the biosynthetic reaction catalyzed by adenylylated glutamine synthetase from *Escherichia coli*. *J Biol Chem* **257:**289–297.

156. Fersht A. 1999. Structure and Mechanism in Protein Science: a Guide to Enzyme Catalysis and Protein Folding. W. H. Freeman, New York, NY.

157. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ. 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protocols* **2:**727–738.

158. Feist AM, Palsson BØ. 2008. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotech* **26:**659–667.

159. Neidhardt FC, Umbarger HE. 1996. Chemical composition of *Escherichia coli*, p 13–16. *In* Neidhardt FC, Curtis R III, Ingraham J, Lin ECC, Low KB, Magasanik B, Reznikoff WS, Riley M, Schaechter M, and Umbarger HE (ed), *Escherichia coli and Salmonella*: *Cellular and Molecular Biology*, 2nd ed., vol. 1. ASM Press, Washington, DC.

160. Bledig SA, Ramseier TM, Saier MH Jr. 1996. Frur mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*. *J Bacteriol* **178:**280–283.

161. Iuchi S, Lin EC. 1988. *arcA* (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc Natl Acad Sci USA* **85:**1888–1892.

162. Sawers G. 1993. Specific transcriptional requirements for positive regulation of the anaerobically inducible *pfl* operon by ArcA and FNR. *Mol Microbiol* **10:**737–747.

163. Park SJ, Chao G, Gunsalus RP. 1997. Aerobic regulation of the *sucABCD* genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream *sdhCDAB* promoter. *J Bacteriol* **179:**4138–4142.

164. Park SJ, Gunsalus RP. 1995. Oxygen, iron, carbon, and superoxide control of the fumarase *fumA* and *fumC* genes of *Escherichia coli*: role of the *arcA*, *fnr*, and *soxR* gene products. *J Bacteriol* **177:**6255–6562.

165. Tseng CP. 1997. Regulation of fumarase (*fumB*) gene expression in *Escherichia coli* in response to oxygen, iron and heme availability: role of the *arcA*, *fur*, and *hemA* gene products. *FEMS Microbiol Lett* **157:**67–72.

166. Tseng CP, Albrecht J, Gunsalus RP. 1996. Effect of microaerophilic cell growth conditions on expression of the aerobic (*cyoABCDE* and *cydAB*) and anaerobic (*narGHJI*, *frdABCD*, and *dmsABC*) respiratory pathway genes in *Escherichia coli*. *J Bacteriol* **178:**1094–1098.

167. Bongaerts J, Zoske S, Weidner U, Unden G. 1995. Transcriptional regulation of the proton translocating NADH dehydrogenase genes (*nuoA-N*) of *Escherichia coli* by electron acceptors, electron donors and gene regulators. *Mol Microbiol* **16:**521–534.

168. Green J, Guest JR. 1994. Regulation of transcription at the ndh promoter of *Escherichia coli* by FNR and novel factors. *Mol Microbiol* **12:**433–444.

169. Pellicer MT, Fernandez C, Badia J, Aguilar J, Lin EC, Baldom L. 1999. Cross-induction of glc and ace operons of *Escherichia coli* attributable to pathway intersection. Characterization of the *glc* promoter. *J Biol Chem* **274:**1745–1752.

170. Davies SJ, Golby P, Omrani D, Broad SA, Harrington VL, Guest JR, Kelly DJ, Andrews SC. 1999. Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. *J Bacteriol* **181:**5624–5635.

171. Okinaka RT, Dobrogosz WJ. 1967. Catabolite repression and pyruvate metabolism in *Escherichia coli*. *J Bacteriol* **93:**1644–1650.

172. Saier MH, Jr. 1998. Multiple mechanisms controlling carbon metabolism in bacteria. *Biotechnol Bioeng* **58:**170–174.

173. Botsford JL, Harman JG. 1992. Cyclic AMP in prokaryotes. *Microbiol Rev* **56:**100–122.

174. Sawers G. 2001. A novel mechanism controls anaerobic and catabolite regulation of the *Escherichia coli tdc* operon. *Mol Microbiol* **39:**1285–1298.

175. Plumbridge J. 1998. Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS. *Mol Microbiol* **27:**369–380.

176. Gui L, Sunnarborg A, LaPorte DC. 1996. Regulated expression of a repressor protein: FadR activates iclR. *J Bacteriol* **178:**4704–4709.

177. Cortay JC, Negre D, Galinier A, Duclos B, Perriere G, Cozzone AJ. 1991. Regulation of the acetate operon in *Escherichia coli*: purification and functional characterization of the IclR repressor. *EMBO J* **10:**675–679.

178. Golby P, Davies S, Kelly DJ, Guest JR, Andrews SC. 1999. Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*. *J Bacteriol* **181:**1238–1248.

179. Tanaka Y, Kimata K, Inada T, Tagami H, Aiba H. 1999. Negative regulation of the *pts* operon by Mlc: mechanism underlying glucose induction in *Escherichia coli*. *Genes Cells* **4:**391–399.

180. Membrillo-Hernandez J, Lin EC. 1999. Regulation of expression of the *adhE* gene, encoding ethanol oxidoreductase in *Escherichia coli*: transcription from a downstream promoter and regulation by *fnr* and RpoS. *J Bacteriol* **181:**7571–7579.

181. Mikulskis A, Aristarkhov A, Lin EC. 1997. Regulation of expression of the ethanol dehydrogenase gene (*adhE*) in *Escherichia coli* by catabolite repressor activator protein Cra. *J Bacteriol* **179:**7129–7134.

182. Negre D, Oudot C, Prost JF, Murakami K, Ishihama A, Cozzone AJ, Cortay JC. 1998. FruR-mediated transcriptional activation at the ppsA promoter of *Escherichia coli*. *J Mol Biol* **276:**355–365.

183. Quail MA, Guest JR. 1995. Purification, characterization and mode of action of PdhR, the transcriptional repressor of the *pdhR-aceEF-lpd* operon of *Escherichia coli*. *Mol Microbiol* **15:**519–529.

184. Lee J, Owens JT, Hwang I, Meares C, Kustu S. 2000. Phosphorylation-induced signal propagation in the response regulator ntrC. *J Bacteriol* **182:**5188–5195.

185. Camarena L, Poggio S, Garcia N, Osorio A. 1998. Transcriptional repression of gdhA in *Escherichia coli* is mediated by the Nac protein. *FEMS Microbiol Lett* **167:**51–56.

186. Harris RM, Webb DC, Howitt SM, Cox GB. 2001. Characterization of PitA and PitB from *Escherichia coli*. *J Bacteriol* **183:**5008–5014.

187. Becker SA, Palsson BØ. 2008. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* **4:**e1000082.

188. Reed JL, Vo TD, Schilling CH, Palsson BØ. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (*i*JR904 GSM/GPR). *Genome Biol* **4:**R54.1–R54.12.

189. Chen L, Vitkup D. 2006. Predicting genes for orphan metabolic activities using phylogenetic profiles. *Genome Biol* **7:**R17.

190. Kharchenko P, Vitkup D, Church GM. 2004. Filling gaps in a metabolic network using expression information. *Bioinformatics* **20**(Suppl. 1)**:**I178–I185.

191. Loh KD, Gyaneshwar P, Markenscoff Papadimitriou E, Fong R, Kim KS, Parales R, Zhou Z, Inwood W, Kustu S. 2006. A previously undescribed pathway for pyrimidine catabolism. *Proc Natl Acad Sci USA* **103:**5114–5119.

192. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:**5691–5702.

193. Osterman A. 2006. A hidden metabolic pathway exposed. *Proc Natl Acad Sci USA* **103:**5637–5638.

194. Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY. 2006. Systems biology as a foundation for genome-scale synthetic biology. *Curr Opin Biotechnol* **17:**488–492.

195. Kim TH, Ren B. 2006. Genome-wide analysis of protein-DNA interactions. *Annu Rev Genomics Hum Genet* **7:**81–102.

196. Wade JT, Struhl K, Busby SJ, Grainger DC. 2007. Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol Microbiol* **65:**21–26.

197. Zhou D, Yang R. 2006. Global analysis of gene transcription regulation in prokaryotes. *Cell Mol Life Sci* **63:**2260–2290.